GRADED GARDEN-PATHS AND THE INFLUENCE OF PHONOLOGICAL CONSTRAINTS ON SYNTACTIC PROCESSING IN REAL-TIME

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Thomas Aaron Farmer

August 2009

GRADED GARDEN-PATHS AND THE INFLUENCE OF PHONOLOGICAL CONSTRAINTS ON SYNTACTIC PROCESSING IN REAL-TIME

Thomas Aaron Farmer, Ph. D.

Cornell University, 2009

Language comprehension, or, the process of extracting intended meaning from an incoming linguistic signal, is a complex task involving an incrementally unfolding interpretation of words within their relevant context. In light of the complexity associated with such a task, it is truly noteworthy that, most of the time, the processes dictating comprehension are carried out rapidly, accurately, and with little apparent conscious effort. The goals of the work provided within this volume were to 1) provide new data that can facilitate discrimination between three different models of on-line syntactic processing, and 2) demonstrate that phonological information is an often-overlooked source of information that can influence language comprehension during both normal reading and the processing of syntactically complex garden-path sentences.

Distinguishing between different accounts of phenomena related to on-line syntactic processing has traditionally been quite difficult. Although several theories of on-line syntactic processing assume the parallel activation of multiple syntactic representations, evidence supporting simultaneous activation has been inconclusive. In Chapters 2 and 3, the continuous and non-ballistic properties of computer mouse movements were examined by recording their streaming x, y coordinates in order to procure evidence regarding parallel versus serial processing. Participants heard structurally ambiguous sentences while viewing scenes with properties either supporting or not supporting the difficult modifier interpretation. The curvatures of the elicited trajectories revealed both an effect of visual context and graded competition between simultaneously active syntactic representations. The results are discussed in

the context of three major groups of theories within the domain of sentence processing, strongly supporting interactive competition-based accounts of syntactic processing over various stage-based accounts.

In relation to the second goal, although it is true that many factors have been demonstrated to affect sentence comprehension, the influence of phonological factors has been all but completely neglected. The aim of the research presented in Chapters 4 and 5 is to demonstrate that phonological typicality, the degree to which the sound properties of an individual word are typical of other words in its lexical category, can influence syntactic processing. First, it is demonstrated that nouns and verbs form separate partially overlapping but coherent clusters in phonological space based on their phonemic properties. Two separate studies demonstrate that phonological typicality affects reading times on target words that occur within a linguistic context heavily biased to contain either a noun or a verb. Finally, a fourth experiment demonstrates that phonological typicality can influence, both on-line and off-line, the processing of syntactic ambiguities arising from the lexical category ambiguity associated with noun/verb homonyms.

Overall, the results of the studies presented here provide strong evidence for a dynamic, highly interactive account of syntactic processing in which any salient and reliable source of information, even information phonological in nature, can aid in the pursuit of accurate, effortless sentence comprehension.

BIOGRAPHICAL SKETCH

Thomas A. Farmer was born and raised in the Washington D. C. metropolitan area. He attended James Madison University in Harrisonburg, VA where he graduated Magna Cum Laude in 2001 with a B. A. in Psychology, and again in 2003 with an M. A. in Psychological Science. After completing four years of his Ph. D. work in the Department of Psychology at Cornell University, he relocated to the Sackler Institute for Developmental Psychobiology at Weill-Cornell Medical College in New York City to complete his final two years of dissertation work. He will graduate from Cornell University in the Summer of 2009 with a Ph. D. in Psychology. His primary research focus is on real-time language processing, with work in the domains of syntactic processing and speech perception. Additionally, he has interests in language-vision interaction, motor systems, bilingualism, and statistical methodology. In the rare event that he is not in the lab, Thomas enjoys traveling, language learning, live music, and happy hour.

ACKNOWLEDGMENTS

First and foremost, I would like to thank the members of my Ph. D. committee: Dr. Morten Christiansen, my graduate advisor, for his patience, generosity, big ideas, and sincere interest in facilitating the development of my academic career, Dr. Michael J. Spivey, for his thoughtful guidance over the course of much of the work presented here, along with his friendship, advice on career development, and the Murder City Devils concert, Dr. Barbara L. Finlay for many provocative conversations about much of the work presented here, her mentorship, and her patently unique sense of humor, and my breadth-member Dr. Melissa J. Ferguson for helping, well, bestow breadth, along with a series of insightful conversation about social cognition. In addition, I would like to thank Shimon Edelman, Michael Goldstein, Jennifer Schwade, David Field, and Tom Gilovich for providing many insights into their respective areas of research, along with my own, and the entire faculty of Psychology from providing a truly stimulating academic environment. Additionally, I would like to thank the faculty members at the Sackler Institute for Developmental Psychobiology in the Department of Psychiatry at Weill-Cornell Medical College, and especially Drs. Jason D. Zevin and Bruce McCandliss, both of whom provided me with a wealth of resources and support. Moreover, I would like to thank the Delores Zohrab Liebmann fund for providing me with three full years of exceedingly generous support.

I would also like to thank the members of Cornell's Cognitive Neuroscience Laboratory, along with the graduate students within whom, or to whom, I either have collaborated, complained, or otherwise confided: Sarah Anderson, Chris Conway , Rick Dale, Lauren Emberson, Mike Hove, Jennifer Misyak, Luca Onnis, Bo Pedersen, Florencia Reali, without whom my experience in graduate school, I believe, would have been substantially weakened.

And of course, I'd like to thank my parents, Tom D. and Cathy Farmer along with Grandma Estep, all to whom I owe a lifetime's worth of gratitude. I'd also like to thank Giuseppe Vezzoli for being pretty awesome and for putting up with me over the past few years. I'm sure that I've inadvertently excluded many important people from this list. In the rare event that any of you actually dig this volume up, I sincerely apologize in advance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction and Overview

Language comprehension, or, the process of extracting intended meaning from an incoming linguistic signal, is a complex task involving an incrementally unfolding interpretation of words within their relevant context. Real-time language comprehension is not a singular, unitary process, but instead can be characterized as a product of multiple perceptual and cognitive systems working in concert under (often times) intense temporal demands. Indeed, during on-line language processing, perceptual encoding of a word must occur, the speech stream must be segmented in order for word recognition and lexical processing to occur, individual words must be concatenated into meaningful units, those units must be combined into larger and larger meaningful units, and all of this must be done, very quickly, while considering complex contextual, semantic, and pragmatic information. In light of the complexity associated with such a task, it is truly noteworthy that, most of the time, the processes dictating comprehension are carried out rapidly, accurately, and with little apparent conscious effort.

The rapidity and accuracy so characteristic of language processing makes it, in many respects, a difficult phenomenon to study. One approach that has proved gainful, however, has been to present participants with sentences that induce some sort of processing difficulty or propensity toward temporary misunderstanding, and then track participant responses to those sentences using experimental paradigms that produce fine-grained dependent measures. For example, consider the famous sentence, “The horse raced past the barn fell.” This sentence, although completely grammatical, is so difficult for the psycholinguistically untrained that they often label it ungrammatical. Such difficulty is a result of the fact that, at least temporarily, multiple possible structural representations exist (see Bever, 1970). In this example, "raced" could either signal the onset of a reduced relative clause, equivalent in meaning to "The horse *that was* raced past the barn …," or, "raced" could be interpreted as the main verb of the sentence, such that the horse is the entity that was willfully racing. If "raced" is for whatever reason initially interpreted as the main verb, then processing difficulty is experienced upon encountering the word *fell* because it requires the less- or non-active reduced relative clause interpretation. This kind of processing difficulty is classically referred to as the garden-path effect. The garden-path effect has proven invaluable in the development of models of on-line syntactic processing, mainly because it is possible to manipulate various linguistic and non-linguistic factors in order to determine what sorts of information influence its presence, and, in the case of the experiments presented here, its magnitude.

In the work presented throughout the remaining chapters of this volume, real-time performance on garden-path sentences is systematically explored in order to 1) aid in the discrimination of three different classes of sentence processing models, and 2) explore the manner in which different information sources can facilitate the comprehension system's ability to arrive at the correct interpretation of a complex sentence. Below, different models of sentence comprehension are briefly presented, and the results of the studies contained in Chapters 2 and 3 are linked to support of an interactive competition-based architecture, at the expense of other accounts. Subsequently, although phonology and syntactic processing are often considered vastly different fields, a basis for phonological contributions to on-line language comprehension is established. The results of the studies contained in Chapters 4 and 5, demonstrating support for an influence of the phonological regularities of individual words on syntactic processing, are briefly reviewed.

Models of Sentence Comprehension

Models of on-line syntactic processing have been developed in order to account for both how garden-paths arise and how the comprehension system recovers from the structural misanalysis when it occurs. Syntax-first models of sentence processing (e.g., Ferreira & Clifton, 1986; Frazier, 1998; Frazier & Clifton, 1996; Frazier & Fodor, 1978) have traditionally proposed that, at a point of syntactic ambiguity, syntactic heuristics (such as “Minimal Attachment” or “Late-Closure”) alone select a single structure to pursue. Under these accounts, then, the initial interpretation of a sentence’s structure is likened to a reflex. Should the initial reflexive selection of one of the multiple possible analyses turn out to be the ultimately correct analysis, given the downstream information contained within the rest of the sentence, then processing proceeds normally, with little detectable evidence of an ambiguity effect. Should the syntactic heuristic select the ultimately incorrect analysis, however, a garden-path effect arises, and a separate reanalysis mechanism is subsequently engaged to “repair” the incorrectly parsed sentence. Importantly, unlike the mechanism responsible for the initial interpretation of the sentence’s structure, which is only guided by syntactic information, the reanalysis mechanism is sensitive to multiple types of information, even information non-syntactic in nature, that can serve to facilitate recovery from the misanalysis. So, models falling into this class are stage-based (initial parse versus reanalysis), modular (non-syntactic information only influences interpretation at a later re-analysis stage), and only allow for one of the possible representations to be active at any given time.

Multiple-constraint based theories (e.g., Green & Mitchell, 2006; McRae, Spivey-Knowlton, & Tanenhaus, 1998; MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, Tanenhaus, & Garnsey, 1994), on the other hand, describe language comprehension as an interactive process whereby all possible syntactic representations are simultaneously partially-active and competing for more activation across time. Unlike the syntax-first models, multiple sources of information, be they syntactic or non-syntactic, integrate immediately to determine the amount of activation provided to each of the competing alternatives. Indeed, factors such as referential context (Altmann, Garnham, & Dennis, 1992; Altmann & Steedman, 1991; Brown, van Berkum, & Hagoort, 2000; Snedeker & Trueswell, 2004; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Spivey & Tanenhaus, 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; van Berkum, Brown, & Hagoort, 1999; van Berkum, Brown, Hagoort, & Zwitserlood, 2003), prosody (Nagel, Shapiro, Tuller, & Nawy, 1996; Snedeker & Trueswell, 2003; Snedeker & Yuan, 2008), and semantic, lexico-frequency, and pragmatic information (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Trueswell, Tanenhaus, & Kello, 1993; Trueswell, 1996; Trueswell, Tanenhaus, & Garnsey, 1994) have all been shown to modulate activation patterns across time, and, in many cases, have even shown a very early influence (to the degree that dependent measures employed are temporally sensitive) of non-syntactic information on the early interpretation of an ambiguity.

In this framework, what feel like garden-path effects are due to the incorrect syntactic alternative winning much of the competition during the early portion of the sentence, and then nonconforming information from the latter portion of the sentence inducing a laborious reversal of that activation pattern. Importantly, the degree to which the incorrect alternative had been winning the competition early on affects the degree to which the reversal of that activation pattern will be protracted and difficult. As a result, one can expect that some garden-path events may be very mild, some moderate, and some extreme, such that a wide variety of sentence-readings should all belong to one population of events with a relatively continuous distribution.

More recently, however, a hybrid model has been proposed. The unrestricted race model (Traxler, Pickering, & Clifton, 1998; van Gompel, Pickering, Pearson, & Liversedge, 2005; van Gompel, Pickering, & Traxler, 2001) follows in the footsteps of constraint-based models in proposing simultaneous integration of multiple constraints from statistical, semantic, and contextual sources. However, rather than ambiguity resolution being based on a temporally dynamic competition process, the unrestricted race model posits an instantaneous probabilistic selection among the weighted alternatives of an ambiguity. Therefore, much like the syntax-first models, it must hypothesize a separate reanalysis mechanism that is responsible for garden-path effects when the initial selected alternative turns out to be syntactically or semantically inappropriate. Thus, the unrestricted race model predicts that sentences with garden-paths and sentences without garden-paths are two separate populations of events (either re-analysis is needed or it is not).

As noted above, syntax-first and constraint-based models of sentence processing can be discriminated based on whether or not the initial processing of a sentence is driven by non-syntactic cues, and indeed, much evidence has accrued in support of early non-syntactic influences. Distinguishing constraint-based accounts from the unrestricted race account of syntactic processing has been substantially more difficult. They both propose that multiple information sources have an early influence on interpretation. The difference between them, however, relates to the degree to which they rely on the activation of one versus multiple syntactic representations at any one time during the comprehension process. Under constraint-based accounts, multiple syntactic representations are simultaneously active, whereas under the unrestricted race account, only one representation, the one that is probabilistically supported over the others, receives 100% of the activation, with the other alternatives being completely discarded. At issue, then, is the degree to which the garden-path effect is a discrete, all-or-none phenomenon whereby only one representation can be active, versus a graded phenomenon in which multiple representations are both partially active.

Goal 1

Although many studies attempting to distinguish syntax-first from constraint-based accounts of syntactic processing have been reported, with most favoring a constraint-based account, few efforts have been made to discriminate between constraint-based and unrestricted race accounts of the data. Accordingly, one goal of the work presented in Chapters 2 (Farmer, Cargill, Hindy, Dale, & Spivey, 2007) and 3 (Farmer, Anderson, & Spivey, 2007) is to attempt such a feat. By employing a relatively new dependent measure, continuous arm-movements indexed by recording the streaming x,y coordinates of a computer mouse, the work in Chapters 2 and 3 raises serious problems for the unrestricted race account. The continuous nature of this dependent measure gives it some advantages, both in terms of the number of data points recorded per second and the smooth curved trajectories produced by the redirection of a movement, over other frequently-used dependent measures. These properties make the measure a good candidate for aiding in the discrimination of the two accounts of interest.

As briefly noted above, these two accounts can be distinguished based on distributional analyses. The unrestricted race account predicts that in conditions where a relatively unbiased ambiguity exists, there should exist a bimodal distribution of some substantial garden-path responses and some non-garden-path responses (either reanalysis is needed or it is not). Under the constraint-based view, however, if both representations of an ambiguity were competing for activation over time, as is the case under a multiple-constraint based account of ambiguity processing, a graded pattern involving some minimal garden-paths, some moderate garden-paths, and some substantial garden-paths is predicted.

The study in Chapter 2 represents an initial test of the degree to which mouse-movements can accurately index the processes underlying syntactic processing, and illuminates some of the advantages of recording arm-movements as opposed to saccades. It is demonstrated that well-established effects from visual-world studies of syntactic processing are replicated, and a set of preliminary distributional analyses of trajectory curvature values in a garden-path condition provides some support against the unrestricted race account. The three studies in Chapter 3 provide a much more detailed examination of the mouse-tracking visual-world paradigm. In the first study, the mouse-movement effects elicited in Chapter 2 are replicated while correcting for a few minor problems in the original set-up and analytic strategy. In the second study, an interactive-activation model based on the central tenants of a constraint-based competition-integration account is employed. In a garden-path condition, it produces a unimodal distribution of responses, helping to bolster the claim that the constraint-based account does in fact predict a unimodal distribution of garden-path effects. Finally, in study three, a non-language study serves to validate the mouse-tracking paradigm. The data in that study demonstrate that in situations where a discrete representational flip occurs, it is highly detectable in the mouse-movement paradigm, thus ruling out the criticism that the continuous nature of mouse movement data masks a fundamentally discrete process.

Phonological Contributions to on-line Syntactic Processing

As evident from the list of constraints (above) that have been shown to exert an early influence on syntactic processing, the comprehension system appears to be sensitive to any source of information that is reliable and salient enough to aid in the comprehension process. It is perhaps somewhat surprising, however, that the role of phonology in on-line syntactic processing has been all but neglected by, more or less, the entire field. Of course, proponents of both constraint-based and syntax-first models have considered the influence of the physical, intonational, and rhythmic properties of the speech stream ("prosody") during on-line processing (e. g. Friederici, 2002; Kelly, 1988; Jun, 2003; Pynte, 1996; Snedeker & Trueswell, 2003), and some attention has been given to the influence of gross-level lexical-based sound constraints, such as lexical stress (Ashby & Clifton, 2005; Gow & Gordon, 1993). Within the domain of *on-line* syntactic processing, however, the degree to which the phonemic properties of individual words can serve as informative cues to various types of syntactic information, and thus be helpful during on-line processing, has not been systematically studied.

Over 15 years ago, Kelly (1992) recognized the lack of any proposed relationship between phonology and syntactic information. He argued that the omission was a result of the superior importance placed upon rules that govern syntactic contingencies, over and above the admittedly probabilistic nature of phonological cues to syntactic forms or categories. The bias bestowed by adherence to Generative Grammar toward the superiority of syntactic rules, however, is probably not the exclusive reason for the lack of research on phonological contributions to syntactic processing and representation. Another factor likely to have contributed to the lack of research on the interface between phonology and syntax is the age-old principle of "the arbitrariness of sign" (Saussure, 1916), which holds that words are simply arbitrary symbols. That is, no aspect of the form of a written word or sound of a spoken word should provide information about its meaning.

The principle of arbitrariness is so ingrained within the field of linguistics that it has been designated a design-feature of language (Hockett, 1960). Although a handful of exceptions to the principle of arbitrariness have been identified (e. g. phonaestemes (Bergen, 2004) and onomatopoeia), it is taken as truth that the sound-meaning relationship is arbitrary (Pinker, 1999). Given the focus on meaning, it may seem as though the principle of the arbitrariness of the sign only has ramifications for the interface (or lack thereof) between phonology and semantics. However, rudimentary forms of syntactic information, such as the grammatical category of a word, are imbued with rudimentary semantic information (e. g. nouns typically denote objects and verbs typically denote actions). As such, the lack of attention paid to sound-meaning relationships may have indirectly hindered exploration into probabilistic relationships between phonological information and various types of syntactic information, such as word class or various syntactic structures.

When systematically investigated, there do seem to be many probabilistic phonological differences between nouns and verbs. Kelly (1992), in a review of phonological cues to nouns versus verbs noted, for example, that nouns tend to have more phonemes and graphemes than verbs, and tend to have stress on the first versus second syllable in disyllabic words. He argued that since there are phonological cues that can probabilistically differentiate nouns from verbs, phonological regularities have the potential to serve as a useful source of information for distinguishing the two word classes. That is, he argued that probabilistic phonological information exists between nouns and verbs, and that people may actually use it when categorizing words from those word classes. Extending the work of Kelly (1992), Monaghan, Christiansen, and Chater (2005) identified 16 potentially useful phonological cues to grammatical category, and showed that those cues, and combinations of them, are significantly diagnostic of nouns versus verbs and open- versus closed-class words.

Goal 2

In light of data supporting the potential usefulness of phonological cues during grammatical category assignment, the second goal of the work contained within this volume is to determine whether phonology can influence the real-time processing of both simple unambiguous sentences and complex ambiguous sentences. That is, the experiments presented in Chapter 4 (Farmer, Christiansen, & Monaghan, 2006) and 5 (Farmer, Monaghan, Misyak, & Christiansen, *submitted*) demonstrate that the sound properties of words can serve as another salient and reliable information source during on-line language comprehension. In Chapter 4, the phonological properties of words were quantified by examining only their phonemic compositions. In study 1, it is demonstrated that nouns tend to sound like other nouns and verbs like other verbs; that is, nouns and verbs form separate coherent, yet partially overlapping, clusters in phonological space. Thus, some words have phonemic properties more typical of their respective lexical class than others. Four experiments are then reported that demonstrate the impact of phonological typicality on the processing of nouns and verbs.

First, it is demonstrated that phonological typicality is a significant predictor of unique variance in word naming latencies. Then, using a self-paced reading methodology, two of the experiments focused on the processing of unambiguous sentences. One experiment involved sentence frames designed to strongly predict that a noun will come next, whereas the frames in the other experiment were created to generate strong expectations for a verb. When the preceding context generated a strong expectation for an upcoming noun, noun-like nouns were read faster than verb-like nouns, and when the context was highly predictive of a verb, verb-like verbs were read faster than noun-like verbs. In a final study, the phonological typicality of a noun/verb homonym is demonstrated to have an effect on which grammatical category gets assigned to the homonym, and as such, which interpretation of an ambiguity that is created by the presence of the homonym gets most heavily pursued.

Moreover, the statistical reliability of the results of the unambiguous noun and verb experiments has recently been questioned. Staub, Grant, Clifton, and Rayner (2009) report a failure to replicate the results of experiments 2 and 3 in Farmer et al. (2006). As such, the data presented in Chapter 5 serve as a response to the failure to replicate, and raise interesting questions regarding the circumstances under which phonological typicality is likely to exert an influence on processing during normal reading.

Summary

The results of the experiments detailed in the remaining pages of this volume make four novel contributions to the field of psycholinguistics and, more broadly, to cognitive science as a whole:

1) arm-movement trajectories can serve as an accurate index of the processes underlying real-time language processing at the level of the sentence;

2) bimodality is not detectable in distributions of garden-path effects, providing problems for accounts of syntactic processing that posit an instantaneous probabilistic selection of only one of multiple interpretations of a structural ambiguity, thus bolstering interactive competition-based accounts of the garden-path phenomenon;

3) the relationship between the phonological form of a word and its meaning may not be entirely arbitrary; instead, subtle probabilistic phonological regularities between nouns and verbs are detectable, providing cues to rudimentary forms of meaning; and,

4) people are sensitive to the probabilistic phonological regularities of nouns and verbs, such that they exert an influence on real-time syntactic processing.

Overall, the results of the studies presented here provide strong evidence for a dynamic, highly interactive account of syntactic processing in which any salient and reliable source of information, even information phonological in nature, can aid in the phenomenon of accurate, effortless sentence comprehension.

REFERENCES

Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language, 31*, 685-712.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247-264.

Altmann, G. T. M., & Steedman, M. J. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191-238.

Ashby, J., & Clifton, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition, 96*, B89-B100.

Bergen, B. K. (2004). The psychological reality of phonaestemes. *Language, 80*, 290-311.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the growth of cognition* (pp. 279-362). New York, NY: Wiley.

Brown, C. M., van Berkum, J. J. A., & Hagoort, P. (2000). Discourse before gender: An event-related potential study on the interplay of semantic and syntactic information during spoken language understanding. *Journal of Psycholinguistic Research, 29*, 53-68.

Farmer, T. A., Anderson, S. E, & Spivey, M. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language, 57*, 570-595.

Farmer, T. A., Cargill, S. A., Hindy, N. C., Dale, R., Spivey, M. J. (2007). Tracking the continuity of language comprehension: Computer mouse-trajectories suggest parallel syntactic processing. *Cognitive Science, 31*, 889-909.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103,* 12203-12208.

Farmer, T. A., Monaghan, P., Misyak, J. B., & Christiansen, M. H. (*submitted*).

Phonological typicality influences sentence processing in predictive contexts.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 248-368.

Frazier, L. (1998). Getting there (Slowly). *Journal of Psycholinguisic Research, 27*, 123-146.

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*, 291-325.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Science, 6*, 78-84.

Gow, D. W., & Gordon, P. C. (1993). Coming to terms with stress: Effects of stress location in sentence processing. *Journal of Psycholinguistic Research, 22*, 545-578.

Green, M., & Mitchell, D. (2006). Absence of real evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language, 55*, 1-17.

Hockett, C. F. (1960). The origin of speech. *Scientific American, 203*, 89-96.

Jun, S. (2003). Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research, 32*, 219-249.

Kamide, Y., Altmann, G. T. M., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133-156.

Kelly, M. (1988). Rhythmic alternations and lexical stress differences in English. *Cognition, 30*, 107-137.

Kelly, M. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99*, 349-364.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676-703.

McRae, K., Spivey-Knowlton, M., & Tanenhaus, M. (1998). Modeling the effects of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 37*, 283-312.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorization. *Cognition, 96*, 143-182.

Nagel, H. N., Shapiro, L. P., Tuller, B., & Nawy, R. (1996). Prosodic influence on the resolution of temporary ambiguity during on-line sentence processing. *Journal of Psycholinguistic Research, 25*, 319-344.

Pinker, S. (1999). *Words and rules: The ingredients of language.* New York: Harper Collins.

Pynte, J. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes, 11*, 165-192.

de Saussure, F. (1916) *Cours de Linguistique Generale*. Paris, France: Paycot.

Snedeker, J., & Trueswell, J. C. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*, 103-130.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49*, 238-299.

Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language, 58,* 574-608.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1521-1543.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of syntactic context on syntactic ambiguity resolution. *Cognitive Psychology, 45*, 447-481.

Staub, A., Grant, M., Clifton, C., Jr., & Rayner, K. (2009). Phonological typicality does not influence fixation durations in normal reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 806-814.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39*, 558-592.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language, 35*, 566-585.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language, 33*, 285-318.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 528-553.

van Berkum, J., Brown, C., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of*

Memory and Language, *41*, 147-182.

van Berkum, J. J. A., Brown, C. M., & Hagoort, P., & Zwitserlood, P. (2003). Event-related potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology, 40*, 235-248.

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language, 52*, 284-307.

van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against constraint-based and two-stage models. *Journal of Memory and Language, 45*, 225-258.

CHAPTER 2

Tracking the Continuity of Language Comprehension: Computer-Mouse Trajectories Suggest Parallel Syntactic Processing*

Sentences such as, “The adolescent hurried through the door tripped” are difficult to process because, at least temporarily, multiple possible structural representations exist (see Bever, 1970). In this example, *hurried* could either signal the onset of a reduced relative clause, equivalent in meaning to *The adolescent who was hurried through the door…*, or, *hurried* could be interpreted as the main verb of the sentence, such that the adolescent is the entity that willfully hurried. If *hurried* is initially interpreted as the main verb, then processing difficulty is experienced upon encountering the word *tripped* because it requires the less- or non-active reduced relative clause interpretation. This kind of processing difficulty is classically referred to as the garden-path effect.

Contemporary accounts of how the comprehension system processes such syntactic ambiguity can be distinguished based on 1) the degree to which they rely on the activation of one versus multiple syntactic representations at any one time during the comprehension process, and 2) the time-frame in which non-syntactic information can constrain interpretation. Syntax-first models (e.g., Ferreira & Clifton, 1986; Frazier & Clifton, 1996) have traditionally proposed that, at a point of syntactic ambiguity, syntactic heuristics alone select a single structure to pursue, and recovery from a misanalysis is achieved via a separate re-analysis mechanism that uses

* Farmer, T. A., Cargill, S. A., Hindy, N., Dale, R., & Spivey, M. (2007). Tracking the continuity of language comprehension: Computer-mouse trajectories suggest parallel syntactic processing. *Cognitive Science, 31,* 889-909. Copyright © 2007 by the Cognitive Science Society. Adapted with permission.

semantic and contextual information. Thus, these models propose that only one representation is active at any given time, and that non-syntactic information only influences interpretation at a later re-analysis stage.

Multiple-constraint based theories (e.g., Green & Mitchell, 2006; McRae, Spivey-Knowlton, & Tanenhaus, 1998; MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, Tanenhaus, & Garnsey, 1994), on the other hand, describe language comprehension as an interactive process whereby all possible syntactic representations are simultaneously partially-active and competing for more activation across time. Unlike the syntax-first models, multiple sources of information, be they syntactic or non-syntactic, integrate *immediately* to determine the amount of activation provided to each of the competing alternatives. In this framework, what feel like garden-path effects are due to the incorrect syntactic alternative winning much of the competition during the early portion of the sentence, and then nonconforming information from the latter portion of the sentence inducing a laborious reversal of that activation pattern. Importantly, the degree to which the incorrect alternative had been winning the competition early on affects the degree to which the reversal of that activation pattern will be protracted and difficult. As a result, one can expect that some garden-path events may be very mild, some moderate, and some extreme, such that a wide variety of sentence-readings should all belong to one population of events with a relatively continuous distribution.

Recently, a sort of hybrid account has emerged that combines certain aspects of each of these theories. The unrestricted race model (Traxler, Pickering, & Clifton, 1998; van Gompel, Pickering, Pearson, & Liversedge, 2005; van Gompel, Pickering, & Traxler, 2001) follows in the footsteps of constraint-based models in proposing simultaneous integration of multiple constraints from statistical, semantic, and contextual sources. However, rather than ambiguity resolution being based on a temporally dynamic competition process, the unrestricted race model posits an instantaneous probabilistic selection among the weighted alternatives of an ambiguity. Therefore, much like the syntax-first models, it must hypothesize a separate reanalysis mechanism that is responsible for garden-path effects when the initial selected alternative turns out to be syntactically or semantically inappropriate. Thus, the unrestricted race model predicts that sentences with garden-paths and sentences without garden-paths are two separate populations of events (either re-analysis is needed or it is not). In other words, in conditions where mean performance is expected to exhibit a garden-path effect, there should exist one of two possible patterns: a) a bimodal distribution of some substantial garden-path responses and some non-garden-path responses, or b) practically all trials exhibiting substantial garden-path effects. A graded pattern involving some minimal garden-paths, some moderate garden-paths, and some substantial garden-paths is not predicted by the Unrestricted Race model.

One source of evidence often used to distinguish between syntax-first and multiple-constraint based accounts of on-line language comprehension comes from eye movements recorded during the comprehension of syntactically ambiguous sentences (like 1a, below) that are presented auditorily while participants are looking at a relevant visual display.

1a) Put the apple on the towel in the box.

1b) Put the apple that's on the towel in the box.

In example (1), the prepositional phrase (PP) *on the towel* creates a syntactic ambiguity in that it could be initially interpreted as a destination (or Goal) for *the apple*, thus attaching to the verb phrase *Put,* or it could be interpreted as a modifier of *the apple*, and thus syntactically attached to that noun phrase. Although corpus analyses have shown that prepositional phrase attachment ambiguities are in general more frequently noun-phrase-attached than verb-phrase-attached (Hindle & Rooth, 1993), in the case of the verb *put* and the ambiguous preposition *with*, there exists a reliable lexically-motivated bias for verb-phrase-attachment (Britt, 1994; Spivey-Knowlton & Sedivy, 1995).

When ambiguous sentences like (1a) are heard in the presence of visual scenes where only one possible referent is present (an apple already on a towel), along with an incorrect destination (an empty towel), and a correct destination (a box), as in the top portion of Figure 2.1, about 50% of the time participants fixate the incorrect destination after hearing the first PP. After the second disambiguating PP is heard, eye movements tend to be re-directed to the correct referent and then to the correct destination. When the unambiguous version of the sentence is heard (1b), participants do not look at the incorrect destination (e.g., the empty towel). The tendency in this one-referent context to look at the incorrect destination until the disambiguating second PP is heard provides evidence of the garden-path effect, and is indicative of initially attaching the ambiguous PP to the verb phrase.

This garden-path effect can, however, be modulated by contextual information contained within the visual scene (Snedeker & Trueswell, 2004; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell, Sekerina, Hill, & Logrip, 1999; see also Knoeferle & Crocker, 2006). When two possible referents (say, an apple on a towel and another apple on a napkin) are present (Figure 2.1, bottom panel) along with an ambiguous sentence like (1a), participants tend to look at the correct referent (the apple on the towel) and move it to the correct destination while rarely, if ever, looking at the incorrect destination. In accordance with previous studies of referential context (e.g., Altmann & Steedman, 1988; Spivey & Tanenhaus, 1998; van Berkum, Brown, & Hagoort, 1999), then, it seems that when two possible referents are present, an expectation is created that they will be discriminated amongst, thus forcing a modifier interpretation of the ambiguous PP. The attenuation of looks to the incorrect destination by the presence of two possible referents, then, is evidence for an early influence of non-syntactic (even non-linguistic) information on the parsing process, and is problematic for traditional syntax-first accounts discussed above.

Although early contextual effects elicited in these and similar visual-world experiments strongly support constraint-based models of human sentence processing over syntax-first models, eye-movement data do not readily afford a clear discrimination between constraint-based and unrestricted race accounts of the data. Within the one-referent context, one might expect that if both possible representations of the ambiguous PP were simultaneously active (as predicted by the constraint-based approaches), participants might, as frequently observed (Tanenhaus et al,. 1995; Spivey et al., 2002), look back and forth between the competitor objects. However, because saccadic eye movements are generally ballistic, they either send the eyes to fixate an object associated with a garden-path interpretation or they don't. The evidence from this paradigm, therefore, is also consistent with the unrestricted race model, where the various constraints are combined immediately, but on any given trial only one syntactic representation is initially pursued. That is, across experimental trials, distributions of eye-movement patterns are almost always bimodal, because the fixations are coded as binomial. There are saccades to locations on the display corresponding to either one of the possible representations, but almost never to a blank region in between those two potential targets. In the following experiment, we examined the dynamics of hand movement in the same sentence comprehension scenario with the goal of determining whether the non-ballistic, continuous nature of computer-mouse trajectories can serve to tease apart these two remaining theoretical accounts.

Experiment 1

Recently, it has been demonstrated that continuous nonlinear trajectories recorded from the streaming *x, y* coordinates of computer-mouse movements can serve as an informative indicator of the cognitive processes underlying spoken-word recognition (Spivey, Grosjean, & Knoblich, 2005), categorization (Dale, Kehoe, & Spivey, 2007), and referential communication (Brennan, 2005). Although individual saccadic eye movements can occasionally show some curvature (Doyle & Walker, 2001; Port & Wurtz, 2003) and some informative variation in landing position (Gold & Shadlen, 2000; Sheliga, Riggio, & Rizzolatti, 1994), individual movements of the arm and hand can show quite dramatic curvature (Goodale, Pélisson, & Prablanc, 1986; Song & Nakayama, 2006; Tipper, Howard, & Jackson, 1997) which can be interpreted as the dynamic blending of two mutually exclusive motor commands (Cisek & Kalaska, 2005; Tipper, Howard, & Houghton, 2000). Additionally, whereas self-paced reading affords 2-3 data points (button-presses) per second, and eye-movement data allow for approximately 3-4 data points (saccades) per second, "mouse-tracking" yields somewhere between 30 and 60 data points per second, depending on the sampling rate of the software used. In light of the ability to record many data points per second, and in light of their ability to curve mid-flight as a result of competition between multiple potential targets, mouse movements have the ability to convey the continuity of processing.

The context and garden-path effects reported in the visual world paradigm are highly replicable when tracking eye movements (Snedeker & Trueswell, 2004; Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999). As such, recording mouse movements in the visual world paradigm can serve as a strong test case by which to evaluate the efficacy of the mouse-tracking procedure for the study of language processing in real-time. If the mouse-tracking technique can produce results from the visual world paradigm commensurate with those obtained by tracking eye movements, we would predict that:

1) averaged trajectories recorded in response to ambiguous sentences in the one-referent context should show significantly more curvature toward the incorrect destination than the averaged trajectories elicited by unambiguous sentences—a pattern corresponding to the garden-path effect, and,
2) the curvature of averaged trajectories in the two-referent condition should not differ statistically between ambiguous and unambiguous sentences, thus demonstrating an influence of referential context on the garden-path effect.

If the influence of referential context is observed, it would provide further evidence against the traditional syntax-first models, but would be consistent with either the constraint-based or the unrestricted race accounts of syntactic processing. The second purpose of this study, then, was to exploit the continuity of the mouse-movement trajectories in order to discriminate between these two remaining theoretical accounts. In order to do so, a measure of curvature magnitude was used to determine the amount of spatial attraction toward the incorrect destination that was exhibited by the ambiguous- and unambiguous-sentence trajectories in the one-referent context. If only one representation were active at any one time, as the unrestricted race account predicts, then the trial-by-trial distribution of trajectory curvatures in the ambiguous-sentence condition should be either: a) bimodal—comprised of highly curved "garden-path" movements and non-curved correct-interpretation movements, or b) uniformly in the more extreme curved range, indicating that almost every trial exhibited a garden-path effect. In contrast, as predicted by the constraint-based approach, if both representations were active and competing simultaneously, one should expect to see a unimodal distribution with a continuous range of non-, somewhat-, and highly-curved trajectories—that is, a gradation of "garden-pathing."

Method

Participants

Forty right-handed native-English speaking undergraduates from Cornell University participated in the study for extra credit in psychology courses. We used only right-handed individuals in order to avoid variability associated with subtle kinematic differences in leftward and rightward movement of the left versus the right arms.

Materials and Procedure

Sixteen experimental items, along with 102 filler sentences, were adapted from Spivey et al. (2002) and digitally recorded. The unambiguous version (1b) of each of the 16 experimental items was recorded first, and then the "that" was removed in order to produce the ambiguous (1a) sentence condition (see Spivey et al., 2002 for details). Each visual context corresponding to the 16 experimental items was varied to produce a one- and two-referent condition. The one-referent visual context (illustrated in Figure 2.1, top) contained the target referent (an apple on a towel), an incorrect destination (a second towel), the correct destination (a box), and a distracter object (a flower). In the two-referent context, all items were the same except that the distracter object was replaced with a second possible referent (such as an apple on a napkin).

Twenty-four filler scenes, designed to accompany filler sentences, were also constructed.

Spoken instructions with a single male voice were recorded using Mac-based digital audio recording software. At the beginning of each sound-file for every item (consisting of a set of three instructions), participants first heard “Place the cursor at the center of the cross.” Then, for the sound-files accompanying scenes that were to be paired with experimental items, the experimental sentence always occurred second, followed by two additional unambiguous filler instructions. For the filler-item scenes corresponding to items without any experimental manipulation, participants heard three scene-appropriate unambiguous instructions. In all cases, two seconds separated the offset of one sentence from the onset of the next sentence within each item.

In critical trials for both the one- and two-referent conditions, the target referent always appeared in the top left corner of the screen, the incorrect destination always appeared in the top right corner of the screen, and the correct destination was always located at the bottom right portion of the screen. The distracter object in the one-referent trials, and the second referent in the two-referent trials, always appeared in the bottom left corner of the screen. Given that the scene layout was held constant across all items in each experimental condition, a left to right movement was always necessary. Although there could exist a systematic bias toward specific locations in the display when moving rightward, this was viewed as unproblematic given that the bias would be held constant across both the ambiguous and unambiguous sentences, which were directly compared in all statistical analyses, for each context. The filler sentences were constructed to prevent participants from detecting any statistical regularities created by the object placements in the experimental trials. In addition to the movement used in the experimental instructions, eleven distinct movements were possible in the visual scene across trials, and an approximately equal number of filler sentences (either eight or ten) were assigned to each of these movements. Therefore, ten sentences required an object in the upper left-hand corner of the display be moved to the upper right corner of the display, eight sentences required an object in the upper left-hand corner of the display be moved to the bottom left-hand corner of the display, and so on.

In each scene, participants saw four to six color images, depending on how many objects were needed for the scene. The images were constructed from pictures of real objects taken by a digital camera and edited in Adobe Photoshop. The visual stimuli subtended an average of 5.96 X 4.35 degrees of visual angle, and were positioned 14.38 degrees diagonally from the central cross. The mouse movements were recorded at an average sampling rate of 40 Hz.

The experimental items were counterbalanced across four presentation lists. Each list contained four instances of each possible condition, but only one version of each sentence frame and corresponding visual context. Two filler sentences were included with the experimental items as described above, and three filler sentences were included with each of 24 distracter scenes. The presentation order was randomized for each participant. Participants were randomly assigned to one of the four presentation lists.

Results

Data Screening and Coding

Mouse movements were recorded during the grab-click, transferal, and drop-click of the referent object in the experimental trials. As a result of the large number of possible trajectory shapes, the *x, y* coordinates for each trajectory from each experimental trial were plotted in order to detect the presence of any aberrant movements. A trajectory was considered valid and submitted to further analysis if it was initiated at the top left quadrant of the display and terminated in the bottom right quadrant, indicating that the correct referent had been picked-up and then placed at the correct destination. This screening procedure resulted in 27 deleted trials, accounting for less than 5% of all experimental trials.

Table 2.1

The errors causing for a trial to be excluded from all analyses, per condition

Error Type	1 Referent, Ambiguous	1 Referent, Unambiguous	2 Referent, Ambiguous	2 Referent, Unambiguous
Target Referent Moved to Incorrect Destination	6	2	1	1
Incorrect Referent Moved to Incorrect Destination	2	0	2	0
Picture Representing a Destination Was Moved	0	0	5	0
Erratic Movement Yielding an Uninterpretable Trajectory	5	1	2	0

The types of errors that resulted in the exclusion of a trial, along with their frequency of occurrence, per condition, are presented in Table 2.1. The most frequent error involved placing the correct referent on the incorrect destination, with no evidence of a corrective movement toward the intended destination. Additionally, errors classified as "erratic" typically contained aberrant movements of the correct referent that can be characterized best as oscillating between rightward movement and leftward movement, with the correct referent either making it eventually to the correct destination or not. A 2 (Context) X 2 (Ambiguity) ANOVA on the number of included trials per condition yielded no significant main effect of context, $F(1, 39)=1.20$, n.s., or two-way interaction, $F(1, 39)=.01$, n.s. There was, however, a significant main effect of ambiguity, $F(1, 39)=9.78$, $p=.003$, $MSE=.134$, with more trajectories included in the unambiguous ($M=7.9$, $SD=.38$) than in the ambiguous ($M=7.42$, $SD=.98$) conditions. The fact that more trials were excluded in the ambiguous conditions is not surprising in light of the increased difficulty associated with the processing of these sentences, and is consistent with error rates in eye-tracking experiments of this type where there are more movement-related errors on ambiguous than on unambiguous trials (Trueswell et al., 1999).

In order to make sure that trajectories in one condition were not initiated (or, that objects were not grabbed) at a systematically different region of the display than in the other conditions, we conducted two 2 (Context) X 2 (Ambiguity) ANOVAs on the x- and y-coordinates, separately. There was no significant main effect or interaction for either the x- or the y-coordinates (all *p*'s n.s.) indicating that, across conditions, the trajectories were initiated at approximately the same location of the display. Subsequently, all analyzable trajectories were "time-normalized" to 101 time-steps by a procedure described in Spivey et al. (2005) and Dale et al. (2007). All trajectories were spatially aligned so that their first recorded point corresponded to x,y coordinates of (0, 0). Although the time-normalized data mirror the general trends evident in raw x- and y-coordinate analyses (see below), they are much more detailed and fine-grained, thus affording more precise information about hand location across time.

Context and Garden-path Effects

The mean trajectories from ambiguous and unambiguous sentences in the one-referent context, illustrated in Figure 2.1 (top), demonstrate that the average ambiguous-sentence trajectory was more curved toward the incorrect destination than the average trajectory elicited by the unambiguous sentences. The point-labels "30^{th}"

through "60^{th}" denote a data-point's corresponding normalized time-step, and they reveal that, in the one-referent context, the average trajectory for the unambiguous sentences traveled to the correct destination much more quickly than did the average trajectory elicited by the ambiguous sentence. Both of these observations support the notion that participants were garden-pathed by the syntactic ambiguity manipulation. Due to the horizontally elongated shape of the overall display, differences in x-coordinates of the mouse movements are somewhat more indicative of velocity differences, and differences in the y-coordinates are more indicative of genuine spatial attraction toward the incorrect destination in the upper right corner. As such, the *t*-tests were conducted across the x-coordinates of each sentence condition, and the y-coordinates of each sentence condition, separately, at each of the 101 time-steps. In order to avoid the increased probability of a Type-1 error associated with multiple *t*-tests, and in keeping with Bootstrap simulations of such multiple *t*-tests on mouse-trajectories (Dale et al., 2007), an observed divergence was not considered significant unless the coordinates between the ambiguous- and unambiguous-sentence trajectories elicited p-values < .05 for at least eight consecutive time-steps. In the one-referent context, two significant divergences were found when comparing the x-coordinates from the ambiguous- and unambiguous-sentence trajectories at each time-step. The comparisons between sentence conditions from time-step 41 to time-step 54 all elicited p-values < .05 (all t's > 2.057, average effect size d=.348). There were also significant differences (p's < .05) in x-coordinates from time-steps 64 to 79 (all t's > 2.05, average effect size d=.347). The y-coordinates at each time-step were compared in the same manner for the ambiguous- and unambiguous-sentence trajectories in the one-referent context. The *t*-tests revealed differences in y-coordinates from time-steps

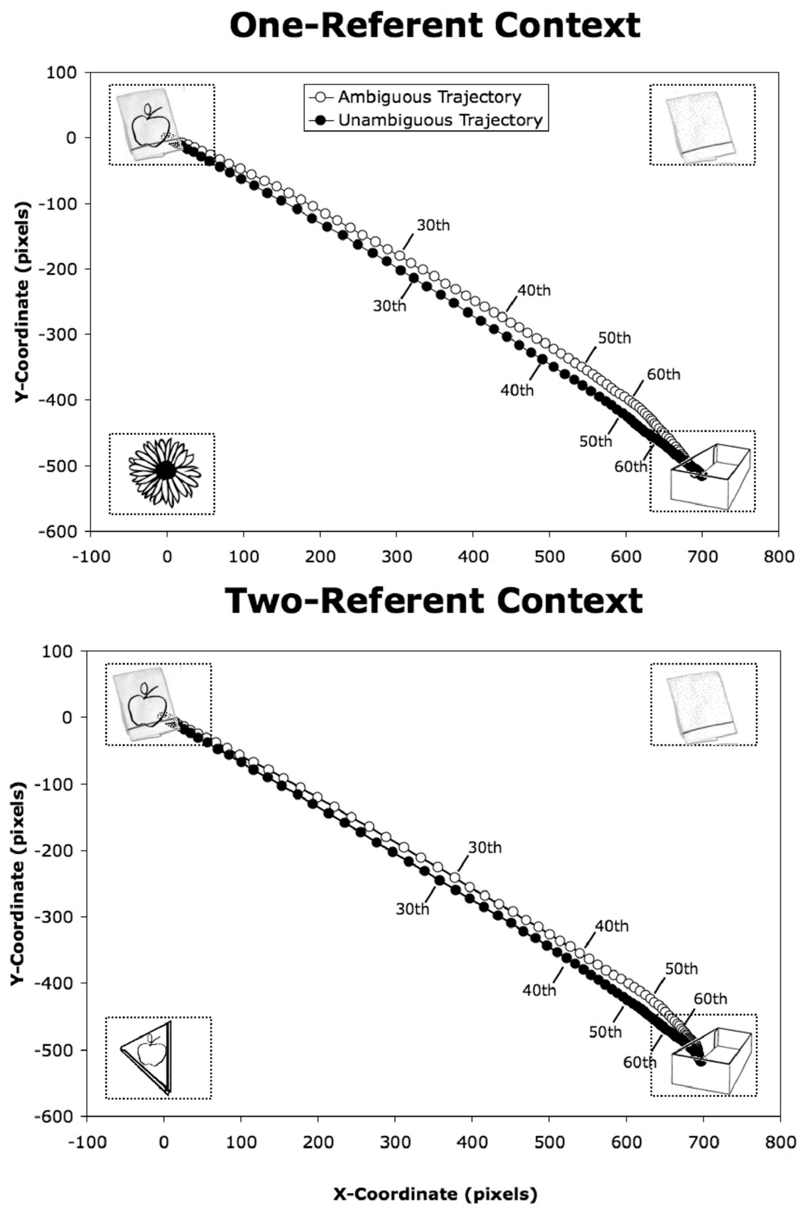

Figure 2.1. An example of a one-referent (top) and a two-referent (bottom) display for the instruction “Put the apple (that’s) on the towel in the box.” The trajectories plotted are the averaged trajectories, per condition, elicited in each context, and the numbers “30th” through “60th” denote a point’s time-step. Due to the horizontally elongated shape of the overall display, differences in x-coordinates of the mouse movements are somewhat more indicative of velocity differences, and differences in the y-coordinates are more indicative of genuine spatial attraction toward the incorrect destination in the upper right corner. Substantial statistically reliable x- and y-coordinate divergence existed between the two sentence conditions in the one-referent context, but both the x- and the y-coordinates for the ambiguous- and unambiguous-sentence trajectories were statistically indistinguishable in the two-referent context.

29 through 82 (all *p*'s $< .05$, all *t*'s > 2.068, average effect size d=.433)[1].

In the two-referent context, the same analyses were conducted on the x- and y-coordinates from the ambiguous- and unambiguous-sentence trajectories at each time-step. For both the x-coordinate and y-coordinate comparisons, importantly, no *t*-test yielded a *p*-value $< .05$ at any of the 101 time-steps.

To address concerns associated with multiple comparisons in the *t*-tests above, and to assess directly the statistical reliability of the Context X Ambiguity interaction, we conducted two separate 2 X 2 X 3 ANOVAs, one for x-coordinates and one for y-coordinates. Based on normalized time-steps, x- and y-coordinates were grouped into three time-bins: 1-33, 34-67, and 68-101, yielding the third independent variable of time segment. The three-way interaction was significant for the x-coordinates, $F(2, 78)=5.06$, $p=.009$, and for the y-coordinates, $F(2, 78)=48.75$, $p < .0005$[2]. As can be observed in Figure 2.1, and as demonstrated by the t-tests above, the effect is especially prevalent among the points comprising time segment two. As such, only the

[1] After examining the trial-by-trial distribution of trajectory curvatures in the one-referent ambiguous-sentence condition (Figure 2.4), one might be concerned that the significant divergences reported are an artifact of the trials in which an extreme garden-path occurred (as indicated by movements all the way to the far upper-right corner of the display). To address this concern, we excluded all trials in the one-referent ambiguous-sentence condition in which the trajectories passed over the incorrect destination before ultimately terminating at the correct destination. Even with these most extreme 5.1% of one-referent trajectories excluded, we still observed significant x-coordinate divergence between the ambiguous- and unambiguous-sentence trajectories from time-steps 39-57 (all *t*'s > 2.02, all *p*'s $< .05$, average d=.36) and 63-82 (all *t*'s > 2.03, all *p*'s $< .05$, average d=.34), and significant y-coordinate divergence from time-steps 39-55 (all *t*'s > 2.06, all *p*'s $< .05$, average d=.35) and from 67-79 (all *t*'s > 2.02, all *p*'s $< .05$, average d=.33).

[2] As per the *t*-test analyses above (see also footnote 1), after excluding the extreme garden-path trials in the one-referent ambiguous-sentence condition, we still observe a significant 3-way interaction for both the x-coordinates, $F(2, 78)=5.07$, $p=.009$, $MSE=2286$, and y-coordinates, $F(2, 78)=3.44$, $p=.037$, $MSE=1291$. Additionally, the Context x Ambiguity interaction at segment two was significant for both the x-coordinates, $F(1, 39)=7.64$, $p=.009$, $MSE=7616$, and marginally for the y-coordinates, $F(1, 39)=3.88$, $p=.056$, $MSE=4987$.

Context X Ambiguity interaction at time segment two will be considered in further detail here.

In this middle time segment, the Context X Ambiguity interaction was significant for both the x-, $F(1, 39)=7.15$, $p=.011$, $MSE=6844$, and the y-coordinates, $F(1, 39)=8.13$, $p=.007$, $MSE=4819$. The means and standard errors for all possible combinations of the independent variables in these x- and y-coordinate analyses appear in Table 2.2. To assess the context effect, we compared each point in the one-referent context to its commensurate point in the two-referent context. For the x-coordinates, there was no difference between coordinates in the one-referent context versus the two-referent context for the unambiguous sentences, $t(39)=.99$, n.s., but there was for the ambiguous sentences, $t(39)=4.14$, $p < .0005$, $d=.655$, with the x-coordinates for the two-referent context being closer to the correct destination. Likewise, for the y-coordinates, there was no difference in average screen location for the unambiguous sentences in the one- versus two-referent context, $t(39)=1.26$, n.s., but there was for the ambiguous sentences, $t(39)=3.71$, $p=.001$, d=.586, with the y-coordinates in the one-referent condition being closer to the top of the display.

In relation to the ambiguity effect for the x-coordinates in this middle time segment, there was no significant difference between ambiguous- and unambiguous-sentence trajectories in the two-referent context, $t(39)=1.65$, n.s., but there was in the one-referent context, $t(39)=2.17$, $p=.036$, $d=.343$, with x-coordinates from the unambiguous-sentence trajectories being closer to the right of the display. For the y-coordinates, there was no significant difference in location between ambiguous- and unambiguous-sentence trajectories in two-referent context, $t(39)=.31$, n.s. However, in the one-referent context, the y-coordinates for the ambiguous-sentence trajectories were significantly closer to the incorrect destination than were the y-coordinates for the unambiguous-sentence trajectories, $t(39)=3.13$, $p=.003$, $d=.495$.

Table 2.2

Means (SE) for the middle-segment ANOVAs

Set	Context	Sentence Type	Mean Coordinate (SE)
X	One Referent	Ambiguous	527.02 (22.47)
		Unambiguous	575.95 (18.26)
	Two Referent	Ambiguous	613.15 (11.70)
		Unambiguous	592.14 (14.01)
Y	One Referent	Ambiguous	-340.06 (19.79)
		Unambiguous	-406.12 (13.81)
	Two Referent	Ambiguous	-416.47 (11.13)
		Unambiguous	-419.95 (9.84)

In order to account for both the x- and y-coordinates in one analysis, we computed the average Euclidean distance at each time-step between corresponding time-steps in the ambiguous- and unambiguous-sentence conditions, per context. Figure 2.2 illustrates that the distance between the ambiguous and unambiguous trajectories in both contexts is similar during the beginning of the trial but then diverges such that the distance between the conditions is considerably larger in the one-referent than in the two-referent context.

Paired-samples *t*-tests, conducted at each time-step as those above, revealed differences in the Euclidean distance between ambiguous and unambiguous sentences in the one versus two-referent context from time-steps 37 through 73, all p's $< .05$ (all t's > 2.11, average effect size d=.459). In Figure 2.1, the averaged ambiguous-sentence trajectory in the one-referent condition is numerically closer to the incorrect destination than its corresponding unambiguous-sentence trajectory across all time-steps. Thus, in the presence of the garden-path effect, it seems clear that there exists more spatial attraction toward the incorrect destination for the ambiguous sentences. It should be noted that the Euclidean distance measure includes both the velocity and spatial attraction effects that cannot be readily delineated given the properties of the scene layout used here. Therefore, in the analyses of the two-referent context, even though the ambiguous- and unambiguous-sentence trajectories are statistically

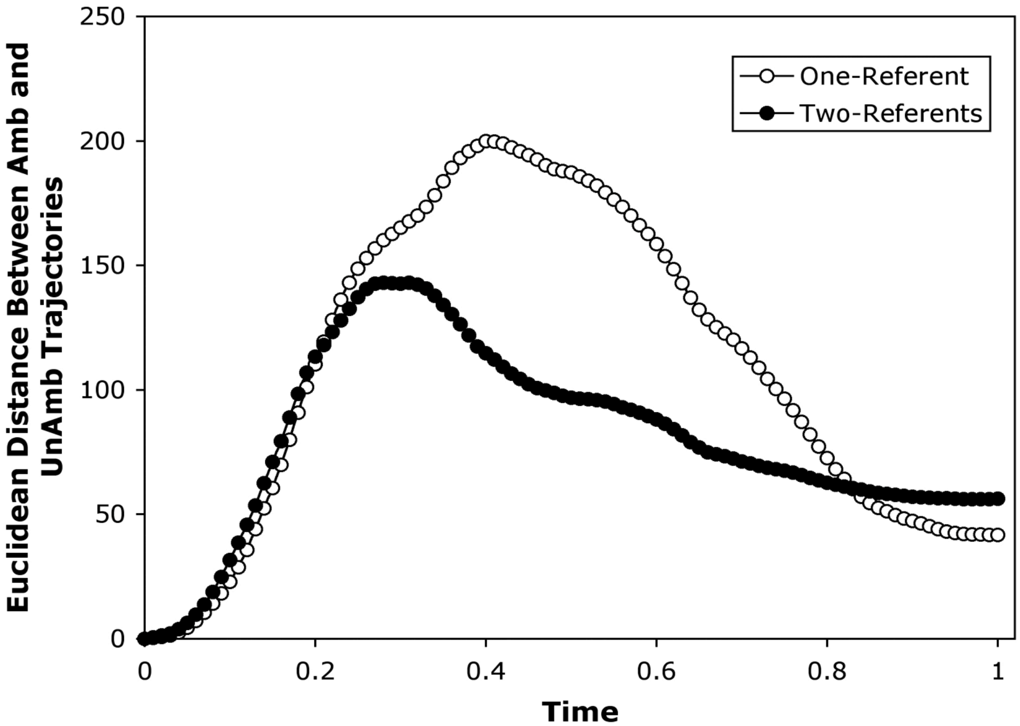

Figure 2.2. The Euclidean distance between the ambiguous- and unambiguous-sentence conditions, per context.

indistinguishable when analyzing x- (more indicative of velocity) and y- (more indicative of spatial attraction toward the competitor) coordinates separately, their combined effects do produce some small coordinate differences between the two sentence conditions. These small coordinate differences in the two-referent condition are, however, largely due to the trajectory in the *ambiguous* condition being faster –

perhaps due to the fact that the unambiguous sentence has a slight delay introduced by the word "that's".

Although analyses of the time-normalized trajectories reveal significant attraction to the incorrect destination in the one-referent ambiguous-sentence condition, two potential criticisms remain. First, it could be argued that the trajectories were initiated, and divergence observed, well after the completion of the spoken sentence, rendering the trajectories, essentially, off-line. Additionally, in light of the velocity difference seen in the one-referent context in Figure 2.1, in which the correct object arrives at the correct destination faster in the unambiguous sentence condition, it could be argued that velocity differences, and not spatial attraction, are driving the statistical significance of the divergence.

To address these concerns, we returned to the raw time-stamps in the trajectories (and their correspondence with portions of the spoken sentences) by examining the average x- and y- coordinates at each of eight different time-bins. The first time bin was composed of the time between the onset of the second (disambiguating) PP up to 250 msec past the onset of that second PP. Each of the following time-bins consisted of consecutive incremental 250 msec intervals, ending with 1750-2000 msec after the onset of disambiguation[3]. As illustrated in Figure 2.3, the trajectories in the ambiguous-sentence condition always lag behind the unambiguous-sentence trajectories in the one-referent condition (x-coordinates), and are always closer to the incorrect destination (y-coordinates). To assess the statistical reliability of these divergence trends, we conducted a *t*-test between the average

[3] Not all trajectories were initiated before the end of the sentence. A participant was included in the analysis if average x- and y-coordinates could be calculated at the time-bin of interest. By time-bin four, notably, most participants were included in the analyses (that is, they had initiated at least 1 trajectory in that condition during the 750-1000 msec time-bin).

ambiguous- and unambiguous- sentence trajectories at each of the 8 time-bins for x- and y-coordinates, separately. To correct for multiple comparisons, the Bonferroni adjustment was used, yielding an adjusted alpha cut-off value of .05/8=.00625.

For the x-coordinates recorded in the one-referent context, average unambiguous- sentence trajectories diverged significantly from average ambiguous-sentence trajectories at time-bins four (750-1000 msec), $t(32)=3.58$, $p=.001$, $d=.624$, and six (1250-1500 msec), $t(38)=2.95$, $p=.005$, $d=.47$ (and marginally significant at bin 5, $t(37)=2.76$, $p=.009$). Thus, we see that in this context, ambiguous sentence trajectories took significantly longer to reach the correct destination than their unambiguous counterparts. More important for the goals of this study, however, we see that there was also significant spatial attraction to the competing incorrect destination. Corresponding analyses of the y-coordinates recorded in the one-referent condition reveal substantial attraction toward the incorrect destination from time-bins four through eight, all t's > 3.20, all p's < .003, average effect size $d=.63$.

Figure 2.3 (bottom panel) illustrates that average y-coordinates from the ambiguous-sentence condition were indeed closer to the top of the screen (y-pixel values closer to zero) than were those of the unambiguous condition trajectories. Additionally, in-line with the time-normalized analyses presented above, none of the eight time-bins in the two-referent context showed the ambiguous- and unambiguous-sentence trajectories significantly diverging for either the x- or the y-coordinates.

Serial versus Parallel Activation

We examined response distributions in the garden-path condition in order to determine whether one or both syntactic representations were active (see Gibson & Pearlmutter, 2000; Lewis, 2000). As an initial attempt to assess whether or not the

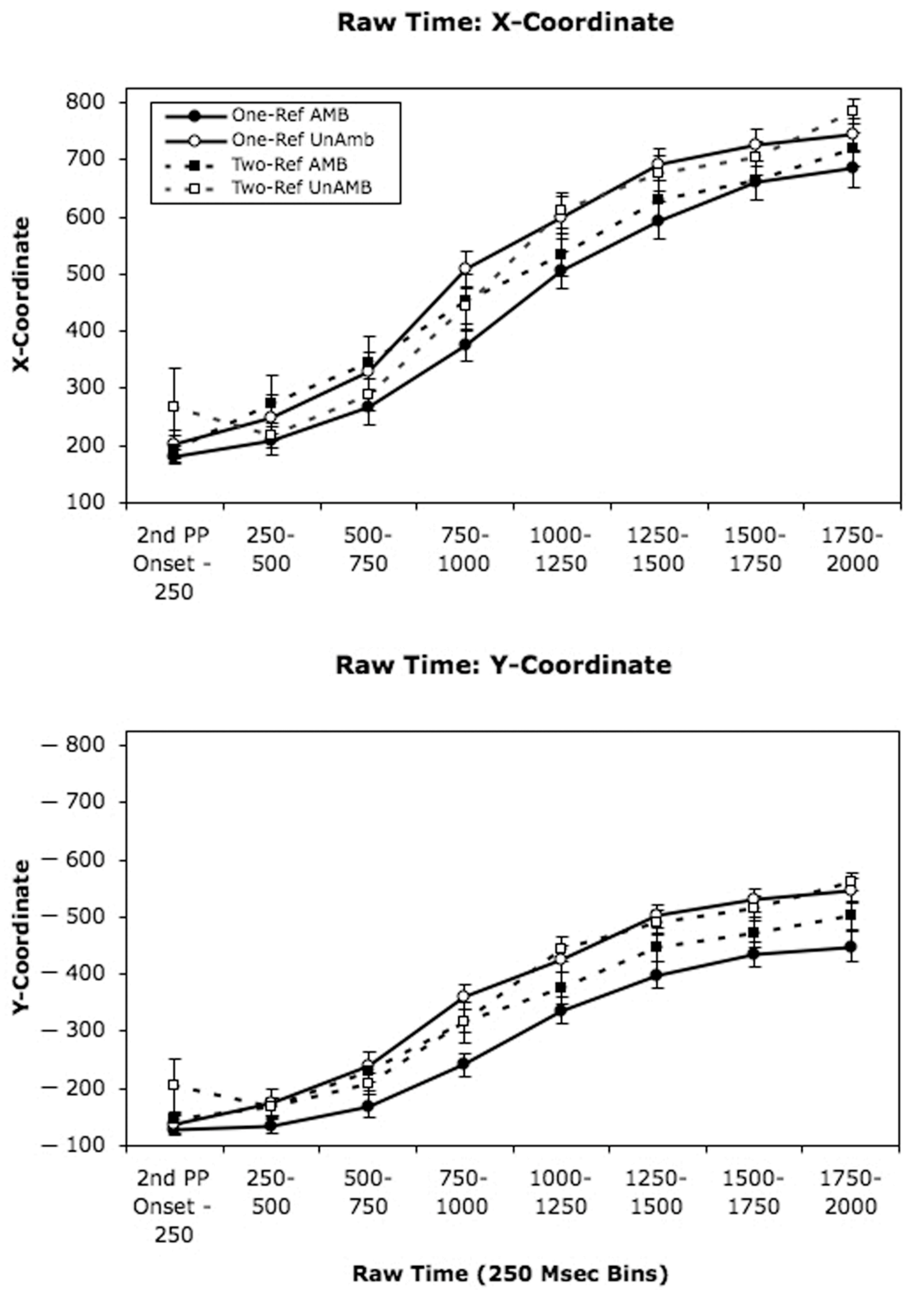

Figure 2.3. In the one-referent context (solid bars), raw non-normalized time bins show x-pixels and y-pixels converging more directly on the correct destination when the instruction is unambiguous than when it is ambiguous. In the two-referent context (dashed bars), this difference between ambiguous and unambiguous instructions is not significant. (Greater positive x values indicate rightward movement and negative y values indicate downward movement.)

distribution of trajectory curvatures in the one-referent ambiguous (garden-path) condition was bimodal (thus indicating only discrete garden-paths and discrete non-garden-paths), we plotted together each of the 146 time-normalized trajectories in that condition, along with a time-normalized reference line from (0, 0) to (700, -500). Figure 2.4 (top panel) illustrates that although there were some extreme garden-path trials and some non-garden-path trials, the majority of the trajectories elicited in this condition fell somewhere in between those two extremes, forming a single population of non-, somewhat-, and highly-curved responses.

To determine whether any bimodality is present in the distribution of responses, we computed the area under the curve on a trial-by-trial basis. First, the straight line from the starting to the ending coordinates of each observed trajectory was normalized to 101 time-steps. Then the total area (in pixels) between that straight line and the observed trajectory was calculated, resulting in an index of trajectory curvature. Area subtending toward the incorrect destination was coded as positive area, and area subtending in the opposite direction from the straight line was coded as negative area. Area of curvature is positively correlated with an alternative measure of curvature, maximum deviation (Atkeson & Hollerbach, 1985), but steady increases in curvature will result in much steeper increases of area than in maximum deviation. Thus, with a much greater range of values in the area measure, the opportunity to observe bimodality in the distribution of curvatures is optimized.

Figure 2.4 (bottom panel) illustrates the shape of the distribution of trajectory curvatures for the one-referent ambiguous-sentence trials. As an index of bimodality, we calculated the bimodality coefficient *b* (SAS Institute, 1989, based on work by Darlington (1970)—see DeCarlo (1997) for a discussion), which has a standard cut-off value of b=.555, with values greater than .555 indicating the presence of bimodality[4]. Although we focus on the one-referent ambiguous response distribution here, Table 2.3 presents the descriptive statistics for each condition's distribution, along with its corresponding bimodality statistic value. The b value for each distribution is less than .555, indicating no presence of bimodality within the distributions. Notably, with regard to the distribution of responses in the one-referent ambiguous-sentence condition, $b < .555$ indicates that the graded spatial attraction effects elicited in this condition came not from two different types of trials but from a single population of trials.

To explore further the modality of the distribution, we compared the area-under-the-curve values in the one-referent ambiguous-sentence condition (where garden-pathing was observed) to the one-referent unambiguous-sentence condition (where no garden-paths were predicted by any of the theories outlined in the introduction), and observed very similar distributional properties. The means are, of course, different, but the standard deviations are nearly identical (SD=121,500 and SD=130,300 for the ambiguous- and unambiguous-sentence conditions, respectively), as are the interquartile ranges (178,110 and 221,470). In fact, when the shapes of the two distributions are compared directly through the Kolmogorov-Smirnov Goodness-of-Fit test, we find that they are not statistically different, p>.1. Distributional characteristics of a population of trials that every theory expects would have a

[4] Caution is warranted when interpreting this cut-off value. A bimodality coefficient b =.555 signals the presence of a uniform distribution whereby all values of X within the distribution have an equal probability of occurring; that is, when the distribution is rectangular, b =.555. Importantly, b does not operate like a p-value, such that values approaching p=.05 are informally treated as indicating the existence of a less statistically reliable result than values much lower than p=.05. Instead, the value for the bimodality coefficient b, typically, must surpass b=.555 before one may infer the presence of *any* noteworthy bimodality.

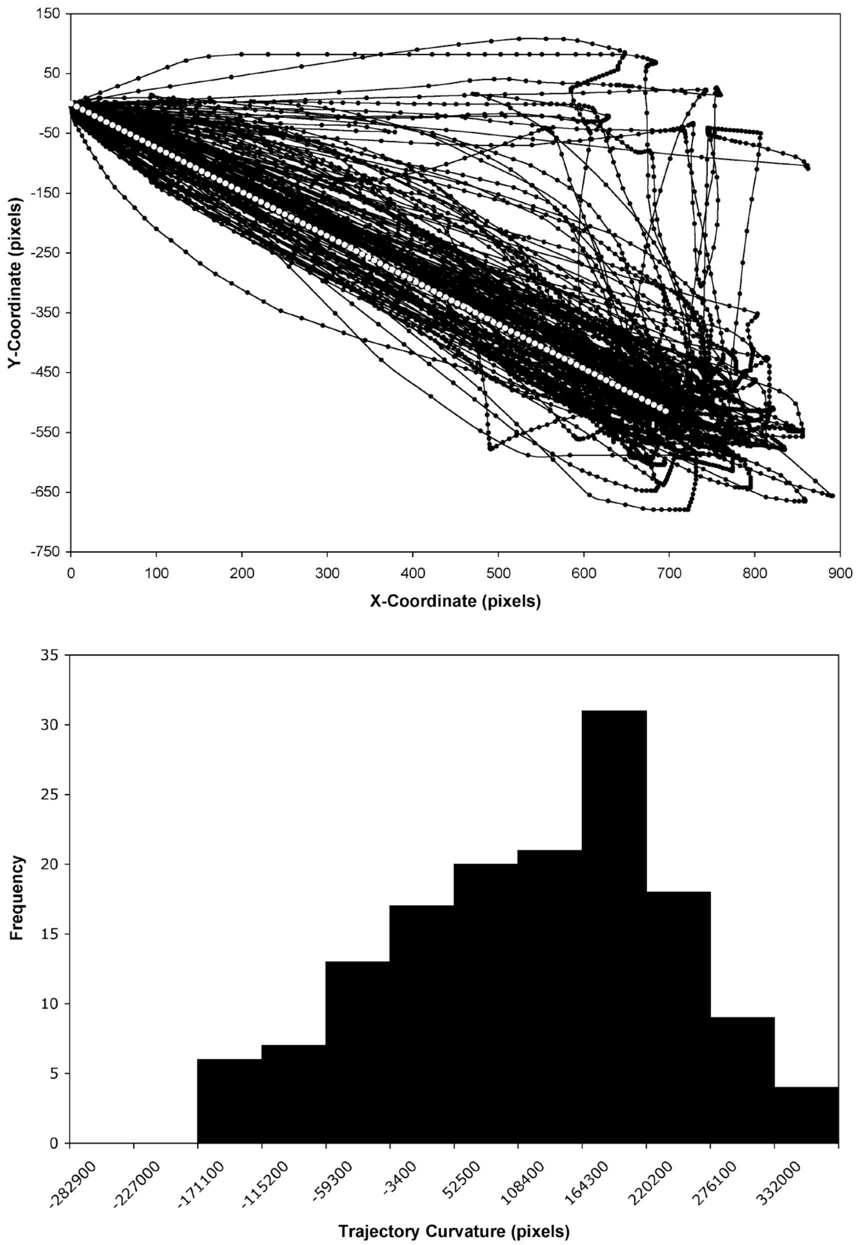

Figure 2.4. Distributions of trajectory curvature in the one-referent ambiguous sentence condition. The top panel illustrates, graphically, that most trajectories curved above a time-normalized reference line (the line of white points), thus illustrating, trial-by-trial, the garden-path effect. The bottom panel illustrates that the distribution of trajectory curvatures is indeed unimodal.

unimodal distribution with no garden-pathing (the unambiguous-sentence condition) and those of a population of trials that should have *substantial* garden-pathing are in fact not distinguishable. This suggests that there is no greater evidence of bimodality in the garden-path condition (where certain theories predict it) than in the unambiguous control condition (where no theory predicts it).

Finally, one might argue that bimodality was not detected (thus, $b < .555$) in the crucial one-referent ambiguous-sentence condition due to a lack of statistical power resulting from the relatively small number of trials in the garden-path distribution. In order to address this concern, we created an artificial distribution with a sample size almost identical to our crucial garden-path distribution by randomly sampling 50% of the trials from the one-referent ambiguous-sentence condition (where garden-pathing was observed) and 50% of the trials from the one-referent unambiguous-sentence condition. This "combination" distribution should produce the response distribution that the unrestricted race account predicts for equibiased syntactically ambiguous sentences—one in which a garden-path would either occur due to the discrete selection of the ultimately incorrect representation, or would not occur, due to the discrete selection of the ultimately correct alternative.

By examining the distributional properties of the area-under-the-curve values produced by the garden-path and non-garden-path trials, together, we can thus determine whether or not the bimodality statistic (b) we used to assess the bimodality of the garden-path distribution (above) is capable of detecting bimodality in a case where the response distribution should clearly be bimodal. Indeed, the bimodality coefficient elicited by this combination distribution (n=151, Skew=-.266, Kurtosis=-1.19) was b=.572. The fact that this bimodal "combination" distribution did elicit a b-

Table 2.3

Statistics necessary for assessing the bimodality of a distribution

Condition	n	Variance	Skewness	Kurtosis	Bimodality (*b*)
1 Referent Ambiguous	147	1.477E+10	-.289	-.535	.429
1 Referent Unambiguous	157	1.699E+10	-.126	-1.141	.529
2 Referents Ambiguous	150	1.629E+10	-.387	-.731	.493
2 Referents Unambiguous	159	1.647E+10	-.545	-.533	.514

value above the absolute cut-off of .555 illustrates that with the sample size used in this study, the bimodality coefficient is capable of detecting bimodality when it should be present (see also Farmer, Anderson, & Spivey, 2007*,* for additional experimental work showing that the mouse-tracking technique can produce bimodal distributions of curvature when they are expected, and that the statistical methods employed here will detect that bimodality).

General Discussion

Converging evidence from the foregoing analyses illustrates that the effects traditionally associated with the visual-world paradigm (Spivey et al., 2002; Tanenhaus et al., 1995) are replicable with the mouse-tracking methodology (see also Magnuson, 2005; Spivey et al., 2005). In the one-referent context, participants' mouse movements in response to the ambiguous sentences curved significantly closer to the top-right of the screen (toward the incorrect destination) than in response to unambiguous sentences. Thus, it would seem that when only one referent was present, the incorrect destination (e.g., the towel) was partially considered relevant, until disambiguating information was processed—a trend corresponding to the garden-path effect associated with this condition. Importantly, any statistically detectable divergence between the x- and y-coordinates of the trajectories in the ambiguous- and unambiguous-sentence conditions was completely absent in the two-referent context, demonstrating that visual context can prevent the syntactic garden-path. The fact that most mouse trajectories began while the speech file was still being heard suggests that the effect of visual context modulating the garden-path took place during early moments of processing the linguistic input, not during a second stage of syntactic reanalysis. Indeed, the time-frame in which significant divergence was observed in the one-referent condition—within one second of the onset of the disambiguating PP—is within the same period of time (relative to the spoken sentence) as when many of the critical fixations of competing objects occur in the visual-world paradigm (Chambers et al., 2004; Spivey et al,. 2002; Tanenhaus et al., 1995; Trueswell et al., 1999).

Additionally, by capitalizing on the continuous, non-linear, and non-ballistic properties of trajectories produced by computer-mouse movements, mouse-tracking has the potential to answer questions that have been difficult to answer with more traditional methodologies. The context effect in the two-referent condition is problematic for syntax-first models of sentence processing, but does not distinguish between constraint-based and unrestricted race accounts. What does distinguish between these two accounts is the gradiency observed in the curvature of the trajectories in the garden-path condition. If the unrestricted race model posits that only one syntactic representation is pursued at any one time, then it must predict that mouse movements in a garden-path condition should initially move either in the direction of the correct destination or in the direction of the incorrect destination (producing either a bimodal distribution or an all-curved distribution). In contrast, since the constraint-based account posits simultaneous graded activation of multiple syntactic alternatives, it predicts that mouse movements can move in directions that are dynamically weighted combinations of the two competing destinations (producing a unimodal distribution of moderate curvatures).

Figure 2.4 shows that although approximately 5% of the trajectories moved all the way to the incorrect destination before changing direction, the vast majority of the trajectories responsible for the mean curvature were unmistakably graded in their degree of spatial attraction toward the incorrect destination. The lack of bimodality in the distribution of trial-by-trial trajectory curvatures suggests that the garden-path effect so frequently associated with this manipulation is not an all-or-none phenomenon. That is, the activation of one structural representation does not forbid simultaneous activation of other possible representations. Instead, the garden-path effect is graded, meaning that although sometimes one syntactic alternative may have greater activation than another, it is also the case that, until disambiguating information is presented, both can be considered in parallel, and the simultaneously active representations may compete for activation over time. Tabor and Hutchins (2004) recently offered evidence of this interpretation. By increasing the length of the region that introduces a garden-path, they showed an increase in the time required to reverse the activation of an incorrect interpretation. Results reveal the gradual commitment to one syntactic interpretation, rather than a discrete selection of one with the immediate dismissal of the others. Their findings, along with the results presented here, appear to strongly support constraint-based accounts of syntactic processing as outlined in the introduction.

More broadly, these results demonstrate that the mouse-tracking technique can be used with tasks that involve complex and interactive displays. We believe that mouse-tracking is a viable method for examining on-line language processing in a wide array of cognitive tasks and across a relatively large age-range. Through a large-scale survey of children's computer use, for example, Calvert, Rideout, Woolard, Barr, and Strouse (2005) found that the mean age at which a child was able to point and click a computer mouse was 3.5 years, and that the mean age of the onset of autonomous computer use was 3.7 years. This observation suggests that experiments employing the mouse-tracking procedure could be feasible with children as young as 3.5-4 years of age, a population for which real-time measures of cognitive processing are often hard to find. Additionally, in light of its accessible, portable, and inexpensive nature, and in light of the replicability of results across the eye- and mouse-tracking methodologies, we believe mouse-tracking can serve as "the poor man's eye-tracker," providing detailed indices of cognitive processing to laboratories that cannot afford expensive eye-tracking equipment. Finally, it is important to note that we do not advocate, or foresee, the usurping of eye-tracking methods in lieu of the advantages of mouse-tracking enumerated here. Instead, we believe that the two techniques can be used in a complementary (even simultaneous) fashion in order to more fully unlock the nature of the complex interactions associated with high-level cognitive processes.

REFERENCES

Altmann, G. T. M., & Steedman, M. J. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191-238.

Atkeson, C. G., & Hollerbach, J. M. (1985). Kinematic features of unrestrained vertical arm movements. *Journal of Neuroscience, 5*, 2318-2330.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the growth of cognition* (pp. 279-362). New York, NY: Wiley.

Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In J. Trueswell & M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-action traditions* (pp. 95-129). Cambridge, MA: MIT Press.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language, 33*, 251-283.

Calvert, S., Rideout, V., Woolard, J., Barr, R., & Strouse, G. (2005). Age, ethnicity, and socioeconomic patterns in early computer use: A national survey. *American Behavioral Scientist, 48,* 590-607.

Chambers, C., Tanenhaus, M. K., & Magnuson, J. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*, 687-696.

Cisek, P., & Kalaska, J. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron, 45*, 801-814.

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing exemplars. *Memory and Cognition*, *35*, 15-28.

Darlington, R. B. (1970). Is kurtosis really “peakedness”? *American Statistician, 24*, 19-22.

DeCarlo, L. T. (1997). On the meaning of kurtosis. *Psychological Methods*, *2*, 292-307.

Doyle, M., & Walker, R. (2001). Curved saccade trajectories: Voluntary and reflexive saccades curve away from irrelevant distractors. *Experimental Brain Research, 139*, 333-344.

Farmer, T. A., Anderson, S. E., & Spivey, M. J. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language, 57*, 570-595.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 248-368.

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

Gibson, E., & Pearlmutter, N. J. (2000). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research, 29*, 231-240.

Gold, J., & Shadlen, M. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature, 404*, 390-394.

Goodale, M., Pélisson, D., & Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature, 320*, 748-750.

Green, M., & Mitchell, D. (2006). Absence of real evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language, 55*, 1-17.

Hindle, D., & Rooth, M. (1993). Structural ambiguity and lexical relations. *Computational Linguistics, 19*, 103-120.

Knoeferle, P. & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science, 30*, 481-529.

Lewis, R. L. (2000). Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research, 29*, 241-248.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676-703.

Magnuson, J. S. (2005). Moving hand reveals dynamics of thought. *Proceedings of the National Academy of Sciences*, *102*, 9995-9996.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the effects of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 37*, 283-312.

Port, N. L., & Wurtz, R. H. (2003). Sequential activity of simultaneously recorded neurons in the superior colliculus during curved saccades. *Journal of Neurophysiology, 90*, 1887-19003.

SAS Institute. (1989). *SAS/STAT user's guide.* Cary, NC: Author.

Sheliga, B. M., Riggio, L., & Rizzolatti, G. (1994). Orienting of attention and eye movements. *Experimental Brain Research, 98*, 507-522.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49*, 238-299.

Song, J., & Nakayama, K. (2006). Role of focal attention on latencies and trajectories of visually guided manual pointing. *Journal of Vision, 6*, 982-99

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proc. National Academy of Sciences, 29*, 10393-10398.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 24, 1521-1543.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of syntactic context on syntactic ambiguity resolution. *Cognitive Psychology, 45*, 447-481.

Spivey-Knowlton, M. J., & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition, 55*, 227-267.

Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 431-450.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

Tipper, S., Howard, L., & Jackson, S. (1997). Selective reaching to grasp: Evidence for distractor interference effects. *Visual Cognition, 4*, 1-38.

Tipper, S. P., Howard, L. A., & Houghton, G. (2000). Behavioral consequences of selection from population codes. In S. Monsell & J. Driver (Eds.), *Attention and Performance XVIII: Control of Cognitive Processes* (pp. 223-245). Cambridge, MA: MIT Press.

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39*, 558-592.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*, 89-134.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language, 33*, 285-318.

van Berkum, J., Brown, C., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*, 147-182.

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*, 284-307.

van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against constraint-based and two-stage models. *Journal of Memory and Language, 45*, 225-258.

CHAPTER 3

Gradiency and Visual Context in Syntactic Garden-Paths

What exactly is a garden-path? About three decades ago, the term was introduced to the psycholinguistic literature in describing what it feels like to have been led astray by syntactic preferences while reading a sentence. The reader reaches some later portion of the sentence, where the syntax and/or the semantics are no longer consistent or sensible with how she's been parsing the sentence up to that point, and she feels as though she has been "led down the garden-path." But is a garden-path due to the discrete computation of a mental representation that turns out to be inappropriate and must then be deleted and replaced by an alternative mental representation—as seen with discrete computing algorithms (e.g., Budiu & Anderson, 2004; Dietrich & Markman, 2003; Newell, 1990)? Or is a garden-path due to multiple partially-active mental representations simultaneously competing with one another—as seen with populations of neurons coalescing into a stable pattern over time (e.g., Desimone & Duncan, 1995; Rolls & Tovee, 1995; for review, see Spivey, 2007)?

Take, for example, Bever's (1970) famous garden-path sentence, "The horse raced past the barn fell." For novices, that sentence often elicits difficulty. They routinely conclude that the sentence is simply ungrammatical. Even once it is explained that the horse is not doing the racing independently, but is instead *being raced* by some unmentioned rider, and that it is the falling that the horse is doing independently, novices often protest that the sentence is still somewhat hard to process. But for the psycholinguistically trained, this sentence has become easy, indeed passé. Perhaps the novel sentences 1a-c can provide a little more freshness. These sentences use the very same syntactic structure as Bever's sentence about the horse. However, rather than causing the reader to exert equivalent amounts of effort in the pursuit of comprehension, they appear to span a continuum of difficulty in the degree to which the reader feels misled. Consider example sentences 1a-c below. Most readers tend to find example 1a to be the most problematic of these three for parsing as a reduced relative clause. In fact, for that sentence, 33 participants produced a mean acceptability rating (on a scale from 1 to 7) of 1.7 (Hare, Tanenhaus, & McRae, 2007). The temptation to interpret the waiter as the agent of the serving event (which would force the reader down the path of constructing a simple main clause, and thus leave the verb *enjoyed* unattachable) is just too strong. In contrast, example 1c seems essentially unproblematic (as suspects are typically the logical objects of detaining events, and thus the reader is encouraged to pursue the relative clause construction). Its mean acceptability rating was 6.8 (Hare et al., 2007). Crucial to our argument here, example 1b may feel somewhere in between these extremes. Indeed, Hare et al.'s participants gave this sentence a mean acceptability rating of 3.6. What does it mean for an individual garden-path effect to feel graded in its intensity?

1a) The waiter served a steak enjoyed it immensely.

1b) The lioness hunted throughout the night was pregnant with cubs.

1c) The suspect detained for questioning was later released.

Syntax-first models of sentence processing have traditionally proposed that, at a point of syntactic ambiguity, syntactic heuristics alone select a single structure to pursue, and recovery from a misanalysis is achieved via a separate re-analysis mechanism that uses semantic and contextual information (e.g., Ferreira & Clifton, 1986; Frazier, 1998; Frazier, & Rayner, 1982; Rayner, Carlson, & Frazier, 1983). Under such circumstances, either the initially-proposed structure was correct and a garden-path is not experienced, or it was incorrect and a garden-path is experienced while the recovery mechanism replaces the previous syntactic structure with a new one. As a result, one should expect that the reading of a sentence where a garden-path takes place and the reading of a sentence where one does not take place each belong to separate populations of events with different distributional properties.

Contrasting with that account, constraint-based approaches have proposed that statistical, semantic, and contextual biases converge the moment they become available in the input to bias multiple parallel syntactic structure alternatives, with a competition process adjudicating among the alternatives encountered at a point of syntactic ambiguity (e.g., Bates & MacWhinney, 1989; Elman, Hare, & McRae, 2004; MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998). In this framework, what feel like garden-path effects are due to the incorrect syntactic alternative winning much of the competition during the early portion of the sentence, and then nonconforming information from the latter portion of the sentence inducing a laborious reversal of that activation pattern. Importantly, the degree to which the incorrect alternative had been winning the competition early on affects the degree to which the reversal of that activation pattern will be protracted and difficult. Thus, one can expect that some garden-path events may be very mild, some moderate, and some extreme, such that a wide variety of sentence-readings should all belong to one population of events with a single continuous distribution. Note that although syntax-first models do not predict the immediate effects of context that constraint-based models predict, they do readily accommodate this kind of gradiency in garden-path magnitude, attributing the varying difficulty to garden-path recovery processes in the re-analysis mechanism (e.g., Bornkessel, McElree, Schlesewsky, & Friederici, 2004; Fodor & Ferreira, 1998).

Recently, a sort of hybrid account has emerged that combines certain aspects of each of these theories. The unrestricted race model of van Gompel and colleagues

(Traxler, Pickering, & Clifton, 1998; van Gompel, Pickering, Pearson, & Liversedge, 2005; van Gompel, Pickering, & Traxler, 2001) follows in the footsteps of constraint-based models in proposing simultaneous integration of multiple graded constraints from statistical, semantic, and contextual sources. However, rather than ambiguity resolution being based on a temporally dynamic competition process, the unrestricted race model posits an instantaneous probabilistic selection among the weighted alternatives of an ambiguity. Therefore, much like the syntax-first models, it must hypothesize a separate reanalysis mechanism that is responsible for garden-path effects when the initial selected alternative turns out to be syntactically or semantically inappropriate. However, unlike syntax-first models, the unrestricted race model should also predict that roughly equi-biased syntactically ambiguous sentences will sometimes elicit a garden-path and sometimes not—thus producing two separate populations of events within the same experimental condition.

There is now a large body of research demonstrating rapid effects of biasing context on syntactic ambiguity resolution in reading (e.g., Altmann & Steedman, 1988; MacDonald et al., 1994; McRae et al., 1998; Spivey & Tanenhaus, 1998; Trueswell, Tanenhaus, & Garnsey, 1994). However, there is also considerable evidence supporting syntax-first models (e.g., Britt, Perfetti, Garrod, & Rayner, 1992; Clifton, Traxler, Mohamed, Williams, Morris, & Rayner, 2003; Ferreira & Clifton, 1986; Rayner, Garrod, & Perfetti, 1992). At the time of this writing, neither the data nor the computational models have succeeded in completely resolving the debate between constraint-based and syntax-first accounts. In the present context, we are more interested in discriminating between constraint-based models and unrestricted race models of language processing in the visual-world paradigm.

Recent research that is specifically aimed at teasing apart these two accounts has focused on showing that reading times for fully ambiguous sentences can be faster than those for disambiguated sentences (Traxler et al., 1998; van Gompel et al., 2001, 2005). This work has relied on the claim that constraint-based models could not accommodate such results, but this turns out to be an erroneous assumption (Green & Mitchell, 2006). Moreover, the reported ambiguity advantage is less apparent when end-of-trial questions encourage a more careful reading mode (Swets, Desmet, Clifton, & Ferreira, 2005).

The crucial distinction that does separate the constraint-based account from the unrestricted race account is the issue of gradiency in the garden-path itself. The constraint-based approach to sentence processing predicts that the full range of garden-path effects should belong to a single population with a unimodal distribution of "garden-path magnitude," whereas the unrestricted race account (with its constrained probabilistic selection of a single syntactic structure) should predict a bimodal distribution of garden-path effects and non-garden-path effects. Since reading times of disambiguating regions in garden-path sentences constitute a continuous variable, one could, in principle, examine the histogram of reading times for garden-path sentences and test it for bimodality. However, one would need thousands of data points to provide an appropriate test of these alternative predictions, and most sentence processing experiments involve about 30-40 participants each providing reading times for about 4-6 sentences in the ambiguous-sentence unsupportive-context condition.

Eye-movement data from the visual-world paradigm (e.g., Altmann & Kamide, 1999; Knoeferle & Crocker, 2006; Snedeker & Trueswell, 2004; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), which are typically interpreted as supporting constraint-based types of models, have not been able to directly address this gradiency issue because the analyses tend to rely on the frequency of discrete fixations of competitor objects in the visual display. That is, since the saccadic eye movement system is largely ballistic and can only either send the eyes to fixate an object associated with a garden-path interpretation or not, the evidence from this paradigm is equally consistent with the unrestricted race model (where the various constraints are combined immediately, but on any given trial the reader is either garden-pathed or not). If the eyes were capable of regularly making substantially curved saccades, then one could imagine a mild garden-path effect manifesting itself as a subtly curved eye movement that went slightly in the direction of the garden-path object and then landed on the correct object. For example, a visual display with a saccade target and a distractor object (or even just the spatial memory of one) can induce a small landing-point deviation of about 8 minutes of arc (away form the distractor), accompanied by some slight curvature of about 8 minutes of arc, in a saccade that spans 7 degrees of visual angle (Doyle & Walker, 2001; Theeuwes, Olivers, & Chizk, 2005; see also Sheliga, Riggio, & Rizzolatti, 1995). However, such subtly curved saccades and slightly deviated landing positions have not been measured in the visual-world paradigm.

What *can* readily show such a curved movement trajectory is a continuous reaching movement of the hand. For example, when participants reach for a target object that shifts location while the arm is in motion, the arm smoothly adjusts its trajectory mid-flight in order to arrive at the target's new location (Goodale, Pélisson, & Prablanc, 1986). Even the mere presence of a distractor object can attract the movement path toward the distractor or, in some cases, repel the movement path away from it (Song & Nakayama, 2006; Tipper, Howard, & Jackson, 1997). Moreover, finger-pointing movements to colored targets show a temporally continuous graded influence from non-conscious color primes smoothly curving their trajectories (Schmidt, 2002).

Spivey, Grosjean, and Knoblich (2005) adapted this technique to record the streaming [x,y] coordinates of continuous computer-mouse movements for studying real-time spoken word recognition. They presented pictures of objects on a computer screen and gave participants pre-recorded spoken instructions such as “Click the carriage,” and “Click the tower.” With the mouse cursor starting at the bottom center of the screen, and the objects displayed in the upper left and right corners, participants generally moved the mouse upward and curving leftward or rightward. Interestingly, when the distractor object’s name shared phonetic features with the target object’s name (e.g., a carrot opposite the carriage, or a towel opposite the tower), the mouse-movement trajectory tended to be conspicuously curved. When the distractor object’s name did not share phonetic features with the target object’s name (e.g., a raccoon opposite the carriage, or a crayon opposite the tower), there was significantly less curvature in the mouse-movement trajectory. These results were interpreted as evidence for parallel partial activation of multiple lexical items competing over time (e.g., Gaskell & Marslen-Wilson, 1999; Luce, Goldinger, Auer, & Vitevitch, 1998; McClelland & Elman, 1986).[5]

With a similar visual display, this kind of continuous competition is also observed in computer-mouse trajectories toward semantic categories for taxonomic classes (e.g., Mammal and Fish), when participants are given atypical animal exemplars to classify (e.g., whale, seal) compared to typical members of those categories (e.g., horse and trout). Dale, Kehoe, and Spivey (2007) found that, in addition to greater overall curvature in trajectories for atypical animals, the very first time-step of mouse-cursor movement revealed a reliable angular difference between typical-animal responses and atypical-animal responses. Thus, in the atypical animal

[5] In eye-movement data from the visual-world paradigm, similar conclusions were made from the averaged curves of proportion of fixations over time of the target object and of the phonologically similar object (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Spivey-Knowlton, 1996).

condition, the very onset of mouse movement was already exhibiting a mixture of spatial attraction toward both the competitor category and the correct category.

Essentially, when two motor commands are being generated at about the same time (Cisek & Kalaska, 2005), the motor movement produced can sometimes be a weighted combination of the two commands, resulting in an action that moves in the direction of a region *in between* the two intended movement destinations (Godijn & Theeuwes, 2002; Gold & Shadlen, 2000). These kinds of results have been interpreted as evidence that the real-time evolution of a perceptual and cognitive decision is coextensive with the real-time evolution of motor commands (Gold & Shadlen, 2001). Moreover, Paninski et al. (2004) have demonstrated a tight link between the continuous dynamics of neuronal population codes in motor cortex and the continuous dynamics of hand movements. Thus, we suggest that, much like eye movements, continuous computer-mouse movements provide a real-time index of the activations of cognitive representations (especially when much of the arm's inertial mass is supported by a table and most of the continuous movement is carried out by wrist and hand muscles). As a result, portions of trajectories that move toward regions in between two visual targets may be indicative of simultaneous partial activation of the two competing cognitive representations that correspond to those targets.

The purpose of the present work is to marshal converging evidence that will speak to this question of whether syntactic garden-path phenomena manifest themselves as a single continuous unimodal distribution of graded investment in the incorrect parse, or as a bimodal distribution of full-investment and of non-investment in the incorrect parse. Our primary piece of evidence (Study 1) comes from novel experimental results that measure real-time language comprehension in a visual context using continuous motor action: computer-mouse movements. Study 2 provides simulation results from a localist attractor-network model of competition between syntactic alternatives that are consistent with the data from Study 1. Study 3 provides a litmus test of the mouse-tracking paradigm's ability to reveal a bimodal distribution of trajectory curvatures when the initial portion of a stimulus sequence temporarily misleads the participant. Overall, results support constraint-based models of sentence processing, where contextual influences are immediately brought to bear in resolving syntactic ambiguities and simultaneous partial activation of the mutually exclusive syntactic alternatives results in continuous gradations of garden-path magnitude.

Experiment 1

In Experiment 1, we exploit the continuous nature of mouse-movement trajectories, in relation to the visual-world paradigm (Chambers, Tanenhaus, & Magnuson, 2004; Snedeker & Trueswell, 2004; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus et al., 1995), in order to examine the nature of graded spatial attraction toward an object corresponding to a competing, but ultimately incorrect, syntactic representation. Preliminary findings in computer-mouse tracking of sentence processing (Farmer, Cargill, Hindy, Dale, & Spivey, in press) have indeed shown effects similar to those found with eye-tracking (Spivey, et al., 2002; Tanenhaus, et al., 1995). When a participant is given an instruction like that in example 2a, and there is only one apple on the screen (one-referent condition), there exists a tendency to drag the apple slightly toward the towel on the way to dropping it in the box (Farmer et al., in press)—matching the high frequency of eye movements to the towel in that condition seen with eye-movement measures (Spivey, et al., 2002; Tanenhaus et al., 1995). This trajectory curvature toward the incorrect destination object (corresponding to an incorrect parse that attaches the first PP to the verb phrase) is essentially absent when the instruction is unambiguous, as in example 2b. (Since there is only one apple in this context, this PP-attachment amounts to an "over-specification" that does not conform well to Gricean maxims. However, this over-specification is held constant across ambiguous and unambiguous conditions, so it should not be responsible for any garden-path effects.) The role of visual/situational context is robustly demonstrated when the display contains two apples (two-referent condition). In this context, participants' trajectories for ambiguous and unambiguous instructions are statistically indistinguishable, indicating that the visual context greatly reduced the activation of the incorrect syntactic structure.

2a) Put the apple on the towel in the box.
2b) Put the apple that's on the towel in the box.

Although those preliminary data are promising, concerns remain with respect to the spatial and temporal comparisons of the x-pixel and y-pixel components of the trajectories in that initial experiment. In order to directly parallel the experimental trials from eye-tracking studies of the visual world, Farmer et al. (in press) placed the correct referent in the top left-hand portion of the screen, and the correct destination at the bottom right-hand portion of the screen. As a result, the diagonal downward-rightward movement conflates velocity toward the correct destination with spatial attraction toward the incorrect destination. That is, changes in x-coordinates do not only indicate velocity, and changes in y-coordinates do not only indicate attraction toward the incorrect destination. Thus, when a significant divergence between average ambiguous- and unambiguous sentence trajectories in the one-referent context is detected, it is unclear whether the divergence is caused by velocity differences between the two sentence conditions, or if genuine spatial attraction toward the incorrect destination is the source the statistically significant difference.

In the present study, a new version of the experiment involves a layout of objects where the correct mouse-movement is a purely horizontal trajectory, traversing from the left side of the screen to the right side, thereby allowing x-pixel analyses to solely reflect velocity. The incorrect destination was at the top center of the display. Therefore, any upward deflection from the horizontal movement plane reflects spatial attraction toward the incorrect object at the top of the screen, independent of velocity. Motivated by constraint-based accounts of sentence processing, we predicted the following: a) that the average ambiguous-sentence trajectory in the one-referent context would curve upward toward the incorrect destination, reliably more so than the average unambiguous-sentence trajectory, and b) that no such divergence would occur in the two-referent context. Additionally, and more important to the main goal of this present paper, we also predicted that the distribution of trajectory-curvatures in the one-referent ambiguous-sentence condition (the garden-path condition) would yield a unimodal distribution, underscoring the notion that the garden-path effect is a graded phenomenon.

Method

Participants

Thirty-three Cornell undergraduates (*M* = 19.97 years, *SD* = 1.2) participated in this experiment for extra-credit in a psychology course. All participants were native English-speakers and all were right-handed.

Materials

The stimuli were presented using Macromedia Director MX, and mouse movements were recorded at an average sampling rate of 40 Hz. The display resolution was set to 1024 x 768. Sixteen experimental sentences, 104 filler sentences, and 40 visual contexts, adapted from Spivey et al. (2002), were combined to form 40 trials, each with one visual scene and three sentences. During an experimental trial, the experimental sentence preceded two filler sentences. In each of the 24 filler trials, three filler sentences were presented.

Spoken instructions were recorded from one male speaker using Mac-based digital-audio recording software. For the experimental sentences, the unambiguous versions (2b) were recorded, and the ambiguous versions (2a) were then created by editing out the word “that’s” from the unambiguous sentences. By creating the target sentences in this way, the resulting ambiguous and unambiguous versions of a sentence frame had nearly identical prosodies, eliminating the influence of prosodic cues to the attachment site of the ambiguous PP within each ambiguous-unambiguous sentence-pair. The sound files were proofed by two independent listeners and were re-recorded if there were any questions regarding accent, prosody, or the quality of the region in the sound file where “that’s” was removed.

Each visual context was composed of 4-6 objects, depending on the set of spoken instructions. A total of 52 objects were used in the visual scenes, and the images of these objects were created using a digital camera and were edited using Adobe Photoshop. Each object subtended an average of 6.5 degrees of visual angle in width by 5 degrees in height. The objects that could be used as a potential destination tended to be slightly larger (9.5 degrees in width by 6 degrees in height) than the potential referent objects (3.5 degrees in width by 4 degrees in height). All objects were again proofed by two other individuals and were reformulated if the image was ambiguous or distracting (due to loud colors or busy patterns).

Objects in each visual scene were presented in a diamond array (Figure 3.1). Objects on the left and right portions of the screen were positioned 14.6 degrees of visual angle from the center of the display, and objects in the upper and lower portions of the screen were positioned 10.6 degrees of visual angle from the center of the screen. Each version of the 16 experimental items required participants to move an object from the left-hand side of the display to the right-hand side of the display. As such, the exact center, in pixels, of each object appearing on the left- and right-handed portions of the screen always had the same y-coordinates, ensuring that no asymmetry existed in the alignment of the objects positioned on the horizontal movement plane.

Each of the sixteen experimental visual contexts was altered in order to produce a one-referent context and a two-referent context. For example, the one-referent visual context corresponding to sentences 2a and 2b consisted of a target referent (an apple on a towel) on the left side of the display, a correct destination (the box) on the right side of the display, an incorrect destination (an empty towel) at the top of the display, and a filler object (the flower) at the bottom of the display, as illustrated in Figure 3.1 (top panel). For the two-referent contexts, the locations of the target referent, the correct destination, and the incorrect destination were the same. However, instead of a distracter item appearing at the bottom center of the display, a second potential referent was included (an apple on a napkin), as illustrated in Figure 3.1 (bottom panel). In both visual contexts, the distance from the target referent to the incorrect destination was 16.1 degrees of visual angle, and to the correct destination was 25.7 degrees of visual angle.

Following Spivey, et al. (2005), the movements required to complete the remaining 104 filler commands were designed to avoid any movement-based statistical regularities. In addition to the movement used in the target commands, eleven distinct movements were possible in the visual context, and an approximately equal number of filler sentences (either eight or ten) were assigned to each of these movements. For example, ten sentences required an object to be moved from the right-hand portion of the display to the bottom of the display, and eight sentences required an object to be moved from the top of the display to the bottom of the display, etc. Across the full set of instructions, we balanced the relative proportions of

PPs attaching to the noun phrase and to the verb phrase, as well as the relative proportions of single-PP and double-PP sentences (for details, see Spivey et al., 2002).

Both levels of the Context variable were crossed with both levels of the Ambiguity variable, yielding four versions of each of the 16 experimental items. Four presentation lists were created, and the four versions of each experimental item were counterbalanced across those four lists such that each list contained four instances of each possible Context x Ambiguity combination, but only one version of each item. Participants were randomly assigned to one of the four possible presentation lists, and the order of filler- and experimental-sentence presentation occurring within a list was randomized per participant.

Procedure

Participants were asked to make themselves comfortable in front of the computer screen, adjusting the mouse and mouse-pad to a location on the right-hand side that suited them. First, participants read brief instructions, and upon signaling to the experimenter that they understood the task, were next presented with three practice trials (similar in form to the filler trials), followed by the experimental task. At the onset of each trial, participants were presented with the whole scene with the addition of a central cross. After a 500 ms preview period, participants heard the initial command, “Place the cursor at the center of the cross.” One second after the offset of this command, the central cross disappeared and the first of the triplet of object-movement instructions began to play. Four seconds separated the offset of the each instruction from the onset of the next. At the end of the third instruction, a “Done” button appeared at the bottom of the screen, which participants clicked to signal that they were ready to move on to the next trial. The entire experiment took approximately 25 minutes.

Results

Data Screening and Coding

Mouse movements were recorded during the grab-click, transferal, and drop-click of the referent object in the experimental trials. As a result of the large number of possible trajectory shapes, the x, y coordinates for each trajectory from each experimental trial were plotted in order to detect the presence of any aberrant movements. A trajectory was considered valid and submitted to further analysis if it was initiated at the center left region of the display and terminated in the center right region, indicating that the correct referent had been picked-up and then eventually placed at the correct destination. This screening procedure resulted in the exclusion of 28 trials, accounting for 5.3% of all experimental trials. A 2 (Context) X 2 (Ambiguity) ANOVA on the number of included trials per condition yielded no significant main effect of context, $F_1(1, 32) = .139$, n.s., $minF'(1, 47) = .09$, n.s., or 2-way interaction, $F1(1, 32) = .162$, n.s, $minF'(1, 23) = .03$, n.s. There was, however, a significant main effect of ambiguity, $F_1(1, 32) = 13.91$, $p < .05$, $MSE = .218$, $minF'(1, 35) = 5.56$, $p = .024$, with more trajectories included per participant in the unambiguous ($M = 7.88$, $SD = .33$) than in the ambiguous ($M = 7.27$, $SD = .80$) conditions. The fact that more trials were excluded in the ambiguous conditions is not surprising in light of the increased difficulty associated with the processing of these sentences. Importantly, a majority of the ambiguous-sentence trajectories that were excluded contained aberrant movements of the correct referent that can be characterized best as oscillating between rightward movement and leftward movement, with the correct referent either making it eventually to the correct destination or not. No participant was excluded from subsequent analyses given that all participants produced at least 13 interpretable trajectories out of the 16 experimental trials ($M = 15.15$, $SD = .80$).

Each analyzable trajectory was time-normalized to 101 time-steps by interpolating the full set of recorded x,y coordinates spanning from its grab-click to its drop-click. All trajectories were then spatially aligned so that their first recorded point corresponded to x, y coordinates of (0, 0). As noted previously, due to the horizontal alignment of the target referent and the correct destination, the x-coordinates of the elicited trajectories are solely indicative of velocity toward the correct destination, and the y-coordinates are solely indicative of spatial attraction toward the incorrect destination. As such, x- and y-coordinates were analyzed separately. Importantly, given the spatial alignment of the trajectories to point (0, 0), y-coordinates falling below the horizontal plane at the center of the screen have negative values, whereas the coordinates recorded above the horizontal plane have positive values (see Figure 3.1).

It is worth noting that mouse movements tend to be initiated slightly later than eye movements. Therefore, there can be some concern regarding exactly when, with respect to the timing of the speech stream, the mouse began to move. In order to investigate this, we recorded the exact millisecond within each sound-file at which each trajectory was initiated. Across the four conditions, approximately 70-75% of the trials had their mice in motion while the speech file was still playing, and 85-95% of mouse movements were initiated within two seconds of the onset of the second PP (up to 750 ms after the end of the sentence). When the duration of the mouse-movement itself is included (about two seconds on average), this temporal range is about the same temporal range over which eye movements are typically analyzed for these kinds of sentences in the visual-world paradigm (Chambers et al., 2004; Spivey et al., 2002). Therefore, we conclude that these mouse-movement data are sufficiently on-line with respect to the delivery of the spoken instructions to provide evidence on par with the existing eye-movement data—corresponding approximately to the time range of the second, third, and fourth eye movements during task completion.

The Context and Garden-path Effects

Figure 3.1 illustrates the average ambiguous-sentence and unambiguous-sentence normalized trajectories in the one-referent (top panel) and two-referent (bottom panel) displays. In the one-referent context, there appears to be both a velocity and a spatial attraction difference between the average ambiguous and unambiguous trajectories. Notably, the unambiguous trajectories appear to arrive at the correct destination more quickly than the ambiguous trajectories, and the average ambiguous-sentence trajectory curves more toward the top of the screen (toward the incorrect destination) than its unambiguous counterpart. Both of these observations support the notion that participants were garden-pathed in the scenes where only one referent was present. In the two-referent scene, however, there is no evidence of spatial attraction when comparing the average ambiguous- and unambiguous-sentence trajectories, indicating an elimination of the garden-path effect by referential context.

To determine whether any divergences observed across the ambiguous- and unambiguous-sentence trajectories in the one- and two-referent contexts were statistically reliable, our initial analysis involved a series of paired-sample *t*-tests. The *t*-tests were conducted across the x-coordinates of each sentence condition, and across the y-coordinates of each sentence condition, separately, per context condition, at each of the 101 time-steps. In order to avoid the increased probability of a Type-1 error associated wit multiple comparisons, and in keeping with Bootstrap simulations of such multiple *t*-tests on mouse-trajectories (Dale et al., 2007), an observed divergence was not considered significant unless the coordinates between the ambiguous- and

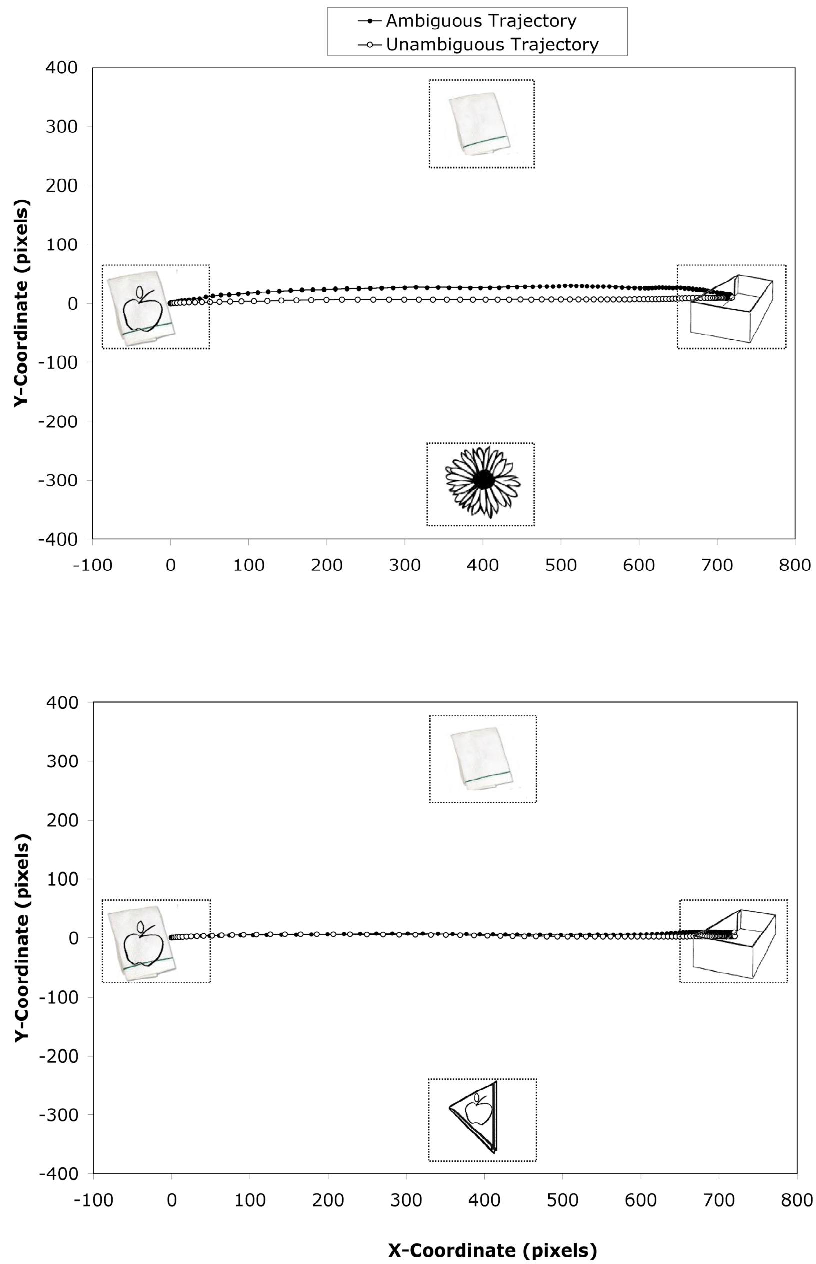

Figure 3.1. A depiction of the typical visual scene, along with the average normalized ambiguous- and unambiguous-sentence trajectories, in the one-referent (top panel) and two-referent (bottom panel) visual-context conditions. There was substantial y-coordinate divergence between the ambiguous- and unambiguous-sentence trajectories in the one-referent context, with ambiguous-sentence trajectories showing more curvature toward the incorrect destination. Additionally, there was absolutely no commensurate divergence in the two-referent context.

unambiguous-sentence trajectories elicited *p*-values < .05 for at least eight consecutive time-steps.

In the one-referent context, the x-coordinates of the average ambiguous- and unambiguous-sentence trajectories differed significantly from time-step 21 all the way to time-step 79, all *t*'s > 2.06, all *p*'s < .05. The average effect size, indicated by Cohen's *d*, was .495, a medium-sized effect in the context of Cohen's benchmarks for effect size (Cohen, 1988). At each of the 59 time-steps in which a significant difference between the ambiguous- and unambiguous-sentence trajectories was observed, the x-coordinates for the unambiguous-sentence trajectories were always higher (that is, they were always further to the right of the screen where the correct destination was located) than they were for the ambiguous-sentence trajectories, indicating a higher velocity of movement, much like that seen with continuous motor responses to high and low frequency words (Abrams & Balota, 1991). Comparisons of the y-coordinates between the ambiguous- and unambiguous-sentence trajectories yielded significant differences from time-step 12 through time-step 72, all *t*'s > 2.043, all *p*'s < .05, average effect size *d* = .502. Thus, very early on in the movement, participants began to exhibit significant spatial attraction toward the incorrect destination. At each of the 51 time-steps in which a significant difference was observed, the y-coordinates were always higher (closer to the location of the incorrect destination at the top of the screen) for the ambiguous-sentence trajectories, suggesting that significant activation of the competing syntactic structure was causing spatial attraction toward the garden-path object.

In the two-referent condition, comparisons of the y-coordinates from the ambiguous- versus unambiguous-sentence trajectories never yielded a single *p*-value < .05 across any of the 101 time-steps. Thus, it appears that in the two-referent context, there was little activation of the competing incorrect interpretation. Interestingly, there was an early but short-lived x-coordinate divergence from time-steps 8 through 21, all t's > 2.06, all p's < .05, average effect size d = .434, with the x-coordinates in the ambiguous-sentence trajectories being closer to the correct destination than the unambiguous-sentence trajectories. This early x-coordinate differential may reflect the fact that in the unambiguous-sentence condition, the delivery of the goal PP in the speech stream is delayed slightly by the word "that's".

To assess directly the statistical reliability of the Context X Ambiguity interaction, we grouped the time-normalized trajectories into four time bins, time-steps 1-25, 26-50, 51-75, and 76-101, yielding a third independent variable of time segment. We then conducted two separate 2 (Context) X 2 (Ambiguity) X 4 (Segment) repeated-measures ANOVAs, one for velocity (x-coordinates) and one for spatial attraction (y-coordinates). The three-way interaction was significant for the x-coordinates, $F_1(3, 96) = 5.30$, p=.002, MSE = 2661, min$F'(3, 135) = 3.10$, p = .029, and for the y-coordinates, $F_1(3, 96) = 2.89$, p = .039, MSE = 128, min$F'(3, 68) = .57$, n.s. As is evident in Figures 3.1 and 3.2, and as demonstrated by the t-tests above, the effect is especially prevalent among the points comprising time-segments two and three. As such, only follow-up comparisons at time-segments two and three are considered in further detail here.

To assess the context effect, we compared each point in the one-referent context to its corresponding point in the two-referent context, in time-segments two and three. The means and standard errors associated with each data-point appear in Figure 3.2, and the confidence intervals reported with each pairwise comparison are the 95% confidence intervals for the mean difference. For the x-coordinates (velocity toward the correct destination), unambiguous sentences showed no difference between the one-referent context and the two-referent context at segment two, $t(32) = .126$,

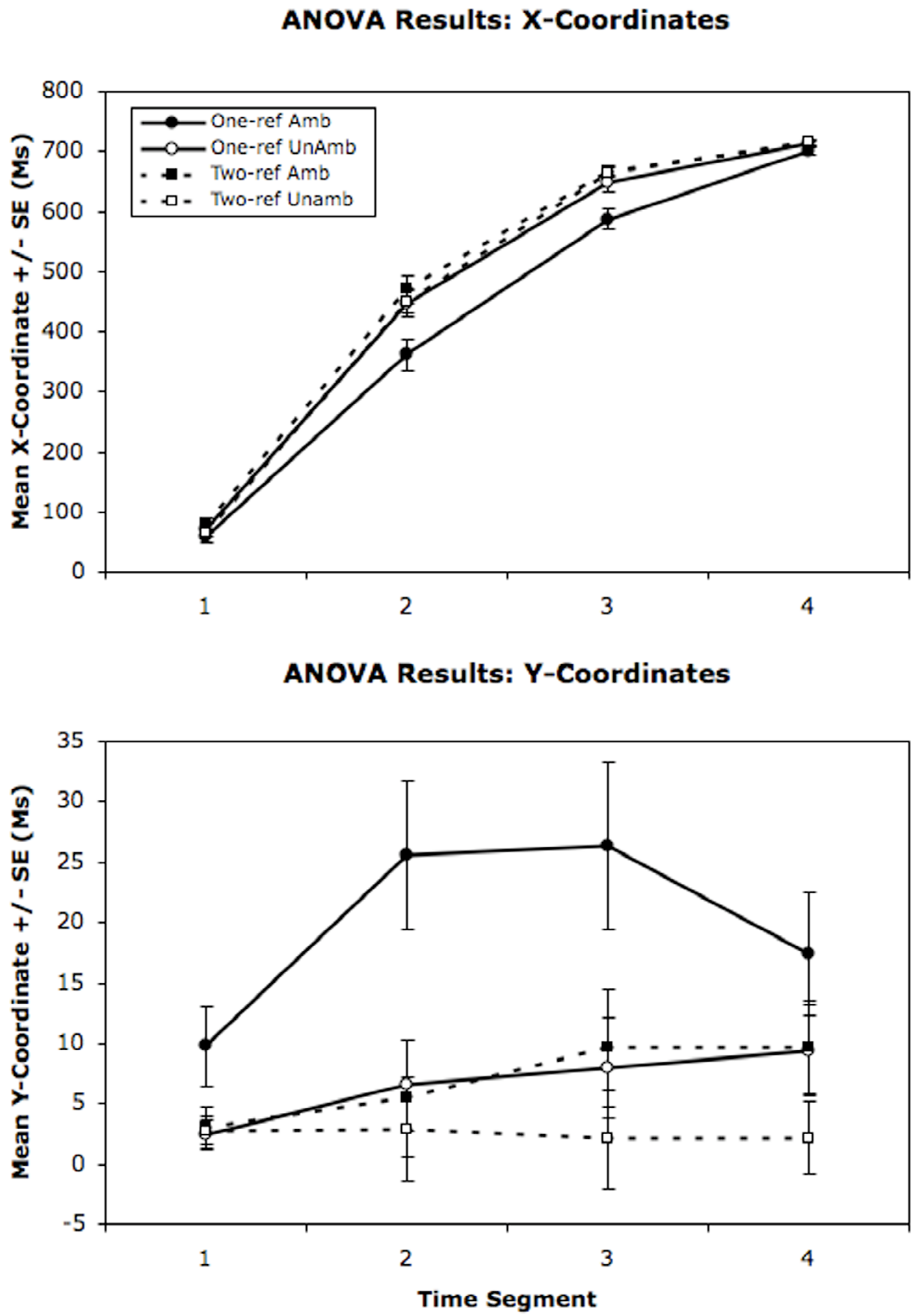

Figure3.2. Average x-coordinate (top panel) and y-coordinate (bottom panel) locations across each of four time-bins. In the one-referent condition, there was significant x- and y-coordinate divergence between the ambiguous- and unambiguous-sentence conditions at segments 2 and 3, but no such divergence in the two-referent context.

n.s., 95% CI = 49.75, or segment three, $t(32) = .867$, n.s., 95% CI = 36.15. However, ambiguous sentences showed a significant velocity difference at segment two, $t(32) = 4.06$, $p < .0005$, $d = .707$, 95% CI = 54.33, and at segment three, $t(32) = 4.63$, $p < .0005$, $d = .798$, 95% CI = 35.20. As illustrated in Figure 3.2 (top panel), the x-coordinates for the ambiguous-sentence trajectories in the two-referent context were closer to the correct destination than they were in the one-referent context at each segment. For the y-coordinates (indicating spatial attraction toward the incorrect destination), unambiguous sentences again showed no difference in average screen location between the one- versus two-referent context at segment 2, $t(32) = .99$, n.s., 95% CI = 7.50, or at segment 3, $t(32) = 1.22$, n.s., 95% CI = 9.93. However, ambiguous sentences did show a spatial attraction difference at segment two, $t(32) = 2.95$, $p = .006$, $d = .513$, 95% CI = 13.89, and at segment three, $t(32) = 2.56$, $p = .015$, $d = .460$, 95% CI = 13.24, with the y-coordinates in the one-referent condition being closer to the incorrect destination at each segment.

In assessing the ambiguity effect, for the x-coordinates, there was no significant difference between ambiguous- and unambiguous-sentence trajectories in the two-referent context at segments two or three, $t(32) = .90$, n.s., 95% CI = 45.44, and $t(32) = .38$, n.s., 95% CI = 24.32, respectively, but there was in the one-referent context at segments two, $t(32) = 3.39$, $p = .002$, $d = .590$, 95% CI = 51.16, and three, $t(32) = 2.96$, $p = .006$, $d = .513$, 95% CI = 41.36, with x-coordinates from the unambiguous-sentence trajectories being closer to the correct destination (Figure 3.2, top panel). For the y-coordinates, there was no significant difference in screen location between ambiguous- and unambiguous-sentence trajectories in the two-referent context at segments two and three, $t(32) = .49$, n.s., 95% CI = 10.74, and $t(32) = 1.21$, n.s., 95% CI = 12.78, respectively. However, in the one-referent context, the y-coordinates for the ambiguous-sentence trajectories were significantly closer to the incorrect destination (top of screen) than were the y-coordinates for the unambiguous-sentence trajectories at segment two, $t(32)=3.53$, $p=.001$, $d=.613$, 95% CI = 10.98, and at segment three, $t(32) = 2.51$, $p = .017$, $d = .423$, n.s., 95% CI = 14.85.

Raw Time Analyses

Given that the trajectories above were time-normalized, the previous analyses do not provide information about when, in relation to the speech-stream, the x- and y-coordinates of the ambiguous-sentence trajectories diverged significantly from those of the unambiguous-sentence trajectories in the one-referent context. In eye-tracking studies that employ a similar manipulation, it has become customary to examine the percentage of looks to the incorrect destination that occur in each of a number of equally-spaced time-bins. Commensurate raw-time analyses are difficult here, however, because trajectory-initiation time varied considerably from trial to trial. A trajectory that was initiated in the one-referent ambiguous-sentence condition 200 milliseconds after the onset of the second PP, for example, is considerably misaligned with a trajectory that was initiated 1000 ms past the second PP onset. Because attraction toward the incorrect destination is not immediate, this misalignment exerts downstream effects whereby the coordinates of trajectories that are initiated in later time-bins (where little attraction has yet to occur) are averaged with the coordinates of earlier-initiated trajectories (where spatial attraction is currently occurring), thus dampening the effect of ambiguity in the one-referent context.

In order to enforce some degree of temporal alignment among the raw-time trajectories, we examined only those trajectories that were initiated before the estimated end-of-sentence time for the longest instruction (2955 ms). By this inclusion criterion, 60% of all experimental trials were included, consisting of 73.11% of the trajectories in the one-referent ambiguous-sentence condition, 52.31% in the one-referent unambiguous condition, 76.47% in the two-referent ambiguous condition, and 36.72% in the two-referent unambiguous condition. To provide a time-course analysis of when during the speech stream the graded spatial attraction toward the garden-path destination emerged, we then examined each of the four 200-ms time-bins occurring between 600 and 1400 ms past the onset of the second PP. This range of time bins corresponds to the central portion of the period of time (during and shortly after the second PP) where previous eye-tracking results have shown the most fixations of the garden-path destination (e.g., Chambers et al., 2004; Spivey et al., 2002; Tanenhaus et al., 1995).

In the one-referent condition, although x-coordinate divergence between the ambiguous- and unambiguous-sentence trajectories did not occur 600-800 ms past the onset of the second PP, there was significant divergence at the 800-1000 ms bin, $t(23) = 2.69$, $p = .013$, $d = .55$, 95% CI = 75.02, the 1000-1200 ms bin, $t(25) = 5.59$, $p < .0005$, $d = 1.08$, 95% CI = 73.75, and the 1200- 1400 ms bin, $t(26) = 7.19$, $p < .0005$, $d = 1.38$, 95% CI = 72.65, with trajectories in the unambiguous-sentence condition being closer to the correct destination on the right than trajectories in the ambiguous-sentence condition. (The df values for each comparison differ because a participant could only be included in the analysis if they had initiated at least one trajectory in both the ambiguous- and unambiguous-sentence conditions at the time-bin of interest.) Marginally significant y-coordinate divergence occurred in the one-referent condition at the 600-800 ms time-bin, $t(18) = 1.97$, $p = .064$, $d = .45$, 95% CI = 17.11, and the 800-1000 ms bin, $t(23) = 1.99$, $p = .059$, $d = .405$, 95% CI = 16.95. Significant divergence did occur in the y-coordinates of the ambiguous- versus the unambiguous-sentence trajectories in the 1000-1200 ms time-bin, $t(25) = 2.31$, $p = .03$, $d = .453$, 95% CI = 18.67. In all cases, the y-coordinates were closer to the incorrect destination at the top of the screen in the ambiguous-sentence condition than in the unambiguous-sentence condition, and there was no significant y-coordinate divergence at the 1200-1400 ms time-bin.

In the two-referent condition, however, both the x- and y-coordinate comparisons between ambiguous- and unambiguous-sentence trajectories showed no significant divergence occurring at any of the four time-bins of interest. Within this subset of the data, then, it seems that the crucial spatial attraction effect (i.e., graded garden-path) is occurring in the one-referent ambiguous-sentence condition at around one second past the onset of the second PP. This result is consistent with what has been seen with similar stimuli in the visual-world paradigm in previous eye-tracking studies (e.g., Chambers et al., 2004; Spivey et al., 2002), where a steady increased level of fixations of the incorrect destination is routinely observed from about 300 to 2000 milliseconds after the onset of the second prepositional phrase.

Distributional Analysis

In addition to demonstrating that mouse-tracking can reveal a visual context's modulation of syntactic garden-path effects, a principal goal of the present study was to examine the distribution of trajectory curvatures in the garden-path condition (one-referent context, ambiguous-sentence). Evidence for bimodality in the distribution of this critical garden-path condition would provide confirmation that some trials involved discrete selection of the incorrect syntactic structure while others involved discrete selection of the correct syntactic structure, as predicted by the unrestricted race account of syntactic ambiguity resolution (Traxler et al., 1998; van Gompel et al., 2001, 2005). In contrast, evidence that the distribution is unimodal would provide support for constraint-based models of sentence processing, where garden-path effects are the continuously graded results of simultaneously partially-active syntactic alternatives competing over time (Elman et al., 2004; Green & Mitchell, 2006; MacDonald et al., 1994; McRae et al., 1998; Tabor & Tanenhaus, 1999). Bimodality in this distribution was initially assessed by visually examining the trial-by-trial overlay of trajectory curvatures from the 119 trials in this garden-path condition. As evident in Figure 3.3, there is a small handful of extreme "garden-path" trials where the trajectory passed over the incorrect destination object before changing direction to move toward the correct destination. However, the vast majority of mouse trajectories that are responsible for moving the mean time-normalized trajectory into the upward bend seen in Figure 3.1 (upper panel) are quite subtle and graded in their curvature. Importantly, there does not appear to be two separate populations of trajectories (e.g., one sizeable group that is essentially straight and horizontal, and another sizeable group that exhibits initial movements toward the incorrect destination followed by corrective turns toward the correct one), as should be predicted by theories that posit immediate probabilistic selection of a single syntactic alternative (Traxler et al., 1998; van Gompel et al., 2001, 2005).

In order to perform statistical tests for bimodality, it was necessary to quantify the magnitude of the garden-path effect within each trial of this one-referent ambiguous-sentence condition. Therefore, we calculated the signed maximum deviation value for each trial by first imposing a straight line from each trajectory's starting point to its endpoint, and then extracting the one point in the observed trajectory with the largest y-coordinate divergence from the straight line. Maximum deviation, in pixels, where the trajectory was above its straight line (tending toward the incorrect destination) was coded as positive, and deviation values produced by trajectories falling below their straight lines were coded as negative. This maximum deviation calculation produces a single value for each trial, indicating the degree of

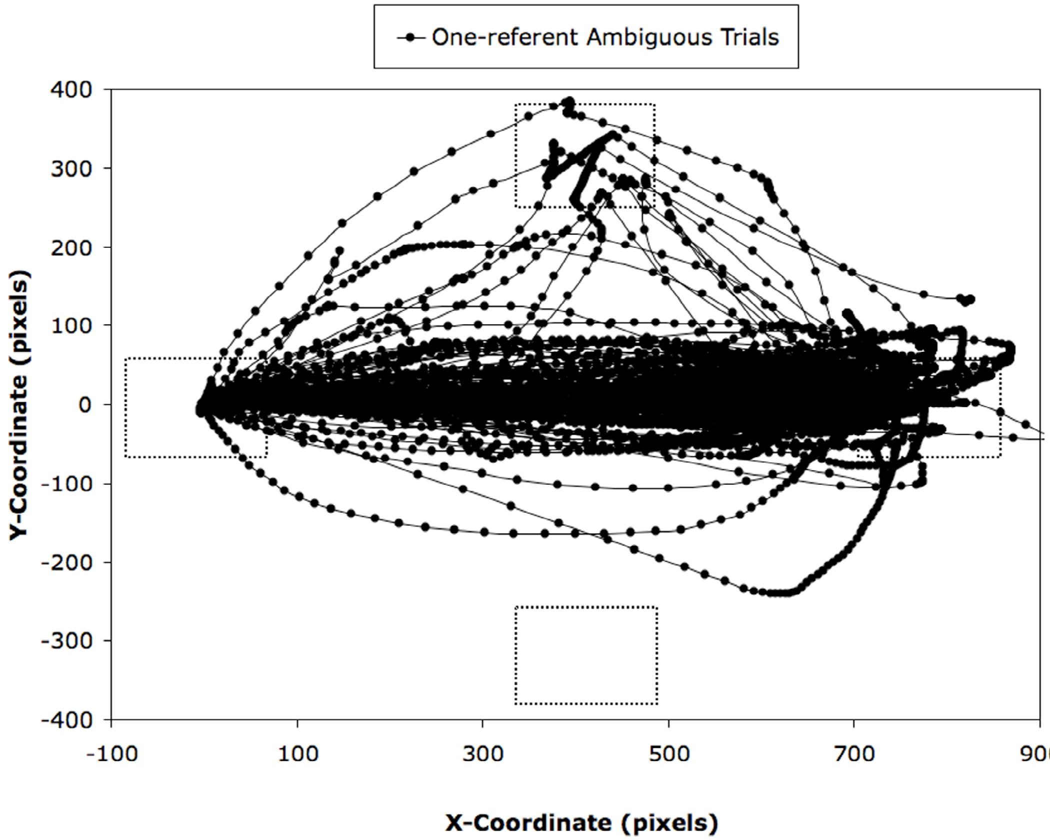

Figure 3.3. The trial-by-trial (n = 119) overlay of trajectories in the one-referent ambiguous-sentence (garden-path) condition. Although a few extreme garden-path trials exist, most trajectories in this condition pass through some intermediate point between the horizontal movement plane and the location of the incorrect destination (top-center) before landing at the correct destination (right-center).

spatial attraction toward the incorrect destination. Using this maximum deviation value, we then plotted a histogram of garden-path-strength values in the 119 trials from this experimental condition and observed that there exists no immediately visible evidence of bimodality in the distribution (Figure 3.4a).

Darlington (1970) noted that an important index of possible bimodality in a distribution is kurtosis (a combination of "peakiness" and "heavy-tailedness" in a distribution). He suggested that a good rule of thumb is that when kurtosis < -1.2, the distribution may have come from two different populations. For non-symmetric

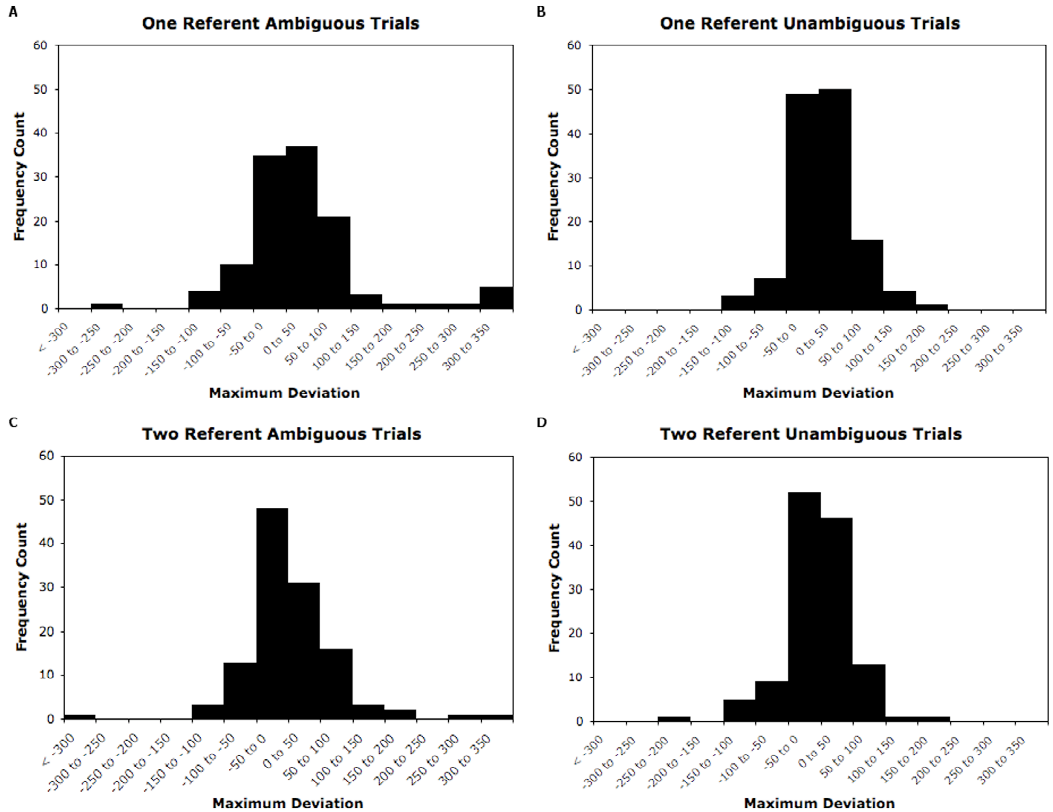

Figure 3.4. The distributions of trajectory curvature magnitudes in the four experimental conditions, calculated as the maximum deviation (in y-pixels) from a straight line. All the distributions are unimodal, indicating that that trajectories elicited in the garden-path condition (panel A) come from one population of garden-path magnitudes, not two.

distributions, later work expanded this measure to include both skewness and kurtosis in the bimodality coefficient (DeCarlo, 1997; SAS Institute, 1989): b = (skewness^2+1) / (kurtosis + (3 * ((n-1)^2) / ((n-2)*(n-3)))), where n equals the number of observations in the distribution of interest. This bimodality coefficient has a standard cut-off value of *b*=.555, with values greater than .555 indicating bimodality in the distribution. Some caution is warranted when interpreting the *b* coefficient in relation to its cut-off value, as statisticians chose this threshold because a uniform (perfectly flat) distribution has a bimodality coefficient of .555. Therefore, distributions whose bimodality coefficients approach this threshold, but are below it, should not necessarily be treated as containing suggestive hints of bimodality, as they are clearly more unimodal than a uniform distribution.

Table 3.1

Maximum deviation statistics for the four distributions of trials in Study 1.

Condition	n	Mean	*SD*	Skewness	Kurtosis	Bimodality (*b*)
1 Referent Ambiguous	119	23.82	91.87	1.40	4.34	.399
1 Referent Unambiguous	130	9.3	50.30	.24	1.26	.244
2 Referents Ambiguous	119	5.24	73.73	.76	6.69	.161
2 Referents Unambiguous	128	-2.5	53.80	-.39	2.59	.203

Table 3.1 presents all the information needed to assess the presence of bimodality within a distribution. Examination of the bimodality coefficient *b* values indicates that no detectable bimodality exists in the distributions of any of the four conditions. Thus, especially in the experimental condition where garden-pathing was observed (one-referent ambiguous-sentence), there do not appear to be two separate populations of garden-path trajectories and non-garden-path trajectories. One might suggest that the few trials falling into the rightmost deviation bin in Figure 3.4a comprise a separate mode within the garden-path trial distribution. It is important to note, however, that even after removing all of the most extreme garden-path trials in

the one-referent ambiguous (n = 7) and unambiguous (n = 1) conditions—the trials where the trajectory actually crossed over the location of the incorrect destination at some point during the trial— there is still substantial evidence for garden-pathing. With those extreme garden-paths removed, significant x-coordinate ambiguous-unambiguous divergence occurred from time-steps 21-79, all t's > 2.07, all p's < .05, average d = .524 (unambiguous always further to the right), and significant y-coordinate divergence still occurred from time-steps 9-43, all t's > 2.04, all p's < .05, average d = .408, with trajectories in the ambiguous-sentence condition being closer to the top of the screen.

Importantly, when comparing the one-referent ambiguous-sentence condition (where substantial garden-pathing is observed) to the one-referent unambiguous-sentence condition (where no garden-paths were expected in the first place), the Kolmogorov-Smirnov Goodness-of-Fit test revealed that the shapes of these two distributions did not differ, p > .1. Hence, we conclude that the distributional properties of a population of trials that should have *no* garden-paths and those of a population of trials that should have *many* garden-paths are not distinguishable, suggesting that there is no greater evidence of bimodality in the garden-path condition (where certain theories predict it) than in the unambiguous control condition (where no theory predicts it). We interpret these results as indicating that there exists a continuum between motor movements elicited by smoothly parsed sentences and those elicited by garden-path sentences.

Discussion

The visual-world paradigm (Tanenhaus et al., 1995; Trueswell, Sekerina, Hill, & Logrip, 1999) allows a behaviorally relevant situational context to impose on-line constraints on real-time sentence comprehension. When adapted for recording the streaming [x,y] coordinates of continuous computer-mouse movements, instead of saccadic eye movements, many of the same findings are observed. The slight loss in immediacy with mouse movements is compensated for by the motor output being much less ballistic than saccadic eye movements, and thereby better able to reveal temporal continuity in the activation changes of mental representations. In our results, substantial evidence from multiple converging analyses supports the notion that both the garden-path effect and the contextual modulation of it were detected by investigating properties of the computer-mouse trajectories recorded in relation to the visual world. In the one-referent context, ambiguous-sentence trajectories took longer to reach the correct destination and were also more curved toward the incorrect destination than were their unambiguous counterparts. This garden-path curvature manifested itself as an average peak deviation (from a straight line) of about 1 degree of visual angle over the course of a 25-degree movement. In contrast, the two-referent context showed very little spatial attraction and no significant difference between the ambiguous- or unambiguous-sentence conditions.

The fact that most mouse trajectories began while the speech file was still being heard suggests that the effect of visual context modulating the garden-path took place during early moments of processing the linguistic input, not during a second stage of syntactic reanalysis. This result is problematic for syntax-first models of sentence processing, but does not distinguish between constraint-based and unrestricted race accounts. What does distinguish between these latter two accounts is the gradiency observed in the curvature of the trajectories in the garden-path condition (one-referent context, ambiguous sentence). If the unrestricted race model posits that only one syntactic representation is pursued at any one time, then one would expect it to predict mouse movements that generally move either in the direction of the correct destination or in the direction of the incorrect destination. In contrast, since the constraint-based account posits simultaneous graded activation of multiple syntactic alternatives, it predicts that mouse movements can move in directions that are essentially weighted combinations of the two competing destinations. Figure 3.3 shows that although 7 of the trajectories moved all the way to the incorrect destination before changing direction, the vast majority of the trajectories responsible for the mean curvature were unmistakably graded in their spatial attraction toward the incorrect destination.

Experiment 2

To further explore this distinction between the predictions of a constraint-based model and those of the unrestricted race account, we designed a competition-integration simulation using the normalized recurrence competition algorithm (Green & Mitchell, 2006; McRae et al., 1998; Spivey & Tanenhaus, 1998). See Figure 3.5. This localist attractor network forces each information source (constraint vector) to provide evidence for the alternatives of an ambiguity in the form of graded support distributed across the alternatives. These constraint vectors send feedforward activation to an integration vector, which then sends feedback to the constraints, gradually biasing them toward the consensually favored alternative. Previous simulations of sentence processing have recorded how long the integration vector takes to settle (with the winning node exceeding an activation threshold), as a measure of reading times. However, the present simulation instead includes a visuomotor constraint vector and converts its dynamic activation patterns into cascaded [x,y] coordinate movements (see Spivey et al., 2005). Therefore, the network actually has three alternatives being competed over: movement toward the correct destination (right side), the incorrect destination (top), and the irrelevant location (bottom), in that order. In the critical target sentences being tested, the first alternative corresponds to attaching a PP to the

NP, the second corresponds to attaching a PP to the VP, and the third corresponds to a viable visuomotor option that is not consistent with the linguistic input on the current trial.

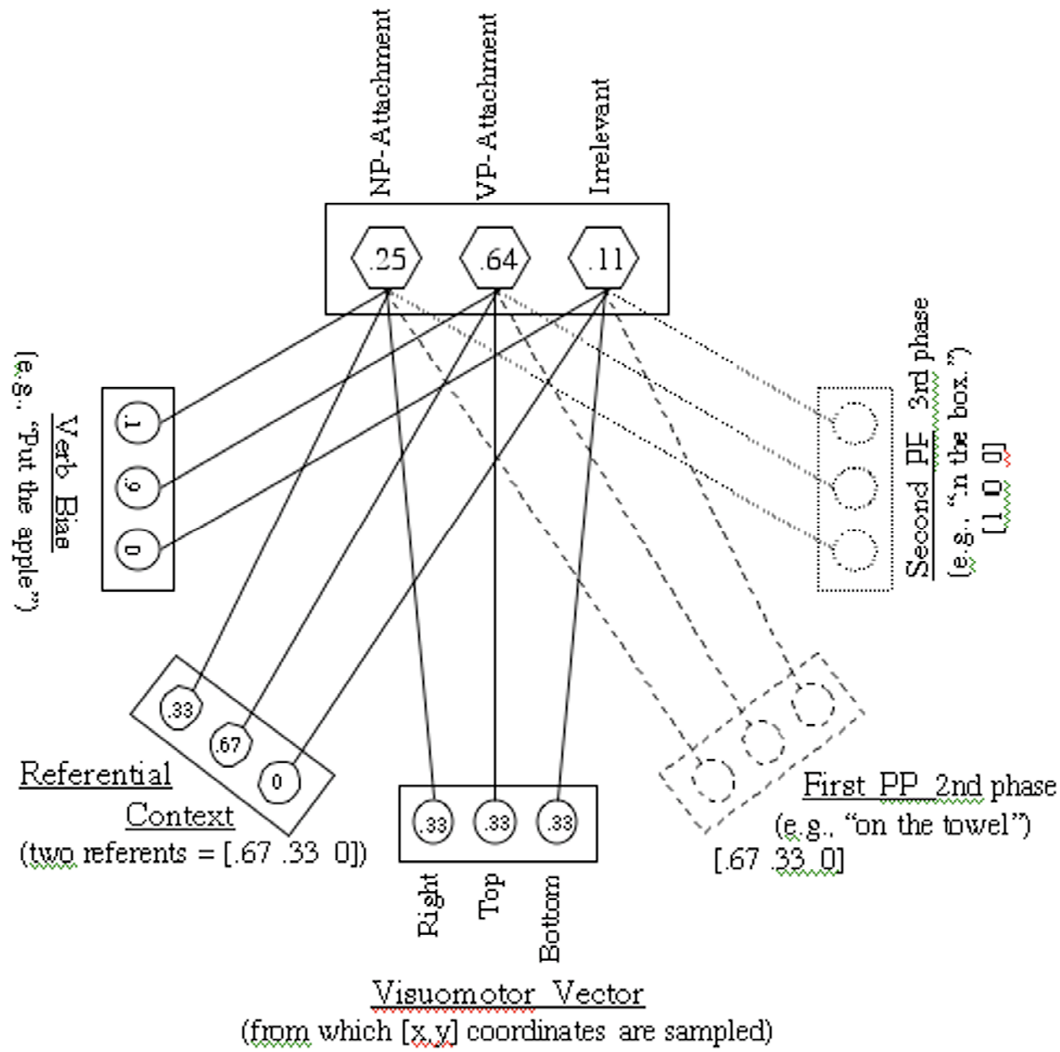

Figure 3.5. In this integration-competition simulation of the one-referent ambiguous-sentence condition, partway through the first time-step, the initial activations of three constraint vectors (verb, context, and visuomotor) produce a weighted average at the integration vector (hexagon nodes). In the second phase o`f competition, hearing the first prepositional phrase activates the First PP vector (dashed lines). In the third phase of competition, hearing the disambiguating prepositional phrase activates the Second PP vector (dotted lines). At every time-step, the activation pattern on the visuomotor vector is converted into a change in [x,y] coordinates by several pixels.

As in previous simulations of garden-path effects, groups of constraint vectors were added to the network in sequence, as the relevant information sources became available in the speech stream (see Figure 3.5). Therefore, the first set of constraint vectors that became active (e.g., for the "Put the apple" phase) were: a) a referential context vector, with the one-referent context moderately supporting verb-phrase-attachment and thus movement toward the incorrect destination [.33 .67 0], and with the two-referent context supporting noun-phrase-attachment and thus movement toward the correct destination [.67 .33 0], and b) a verb-bias vector coding for the fact that "put" has a strong preference to attach any PP to itself (Britt, 1994), supporting movement toward the incorrect destination [.1 .9 0], and c) a visuomotor vector for guiding x,y movements that starts out with uniform random activations between 0 and 1 for each of the three destinations [rand rand rand] to reflect a participant's unpredictable anticipation of where the next instruction might lead them. (This initial random activation pattern in the visuomotor vector is the only non-deterministic aspect of these simulations.) In the ambiguous condition, this first phase of constraints was allowed to compete for three time-steps. In the unambiguous condition, this phase lasted five time-steps (to allow for the duration of the additional word "that's") and also included a constraint vector for hearing the disambiguating "that's" [1 0 0]. The weight for each constraint in this first phase was 1/n, where n is the number of constraints (Spivey & Tanenhaus, 1998). In the second phase, activation patterns from the previous phase carried over and a constraint for the first prepositional phrase ("on the towel") was added, consisting of a moderate statistical bias toward NP-attachment [.67 .33 0], based on the corpus analysis by Hindle and Rooth (1993). As in previous competition-integration simulations, this newly added constraint was given a weight of 0.5, and weights for the other constraints were halved. After five time-steps (for the duration of that phrase), the activation patterns were carried over to the third phase of speech delivery, adding a constraint for the second PP (e.g., "in the box.") with discrete support for noun-phrase-attachment and movement toward the correct destination [1 0 0]. As before, this constraint was given a weight of 0.5, and the weights of the others were halved.

At each time-step, the normalized recurrence competition algorithm combines the constraint vector activations in a weighted average to compute the integration vector's activation pattern, which then returns cumulative multiplicative feedback to the constraint vectors (for details, see Green & Mitchell, 2006; McRae et al., 1998). The feedback from an integration node to a constraint node multiplies the weighted constraint activation that had traveled up that connection by the net activation of that integration node, and adds that product to the constraint node's current activation. Every new time-step begins with the constraint vectors each re-normalizing themselves to sum to 1.0 (applying a form of implicit competition), before being averaged to re-compute the integration vector's new activation pattern.

A key property of this competition algorithm is that the integration and feedback process facilitates a kind of indirect crosstalk whereby the consensus of bias among the majority of constraint vectors can sway any equibiased constraint vectors to follow suit. Recall that the visuomotor constraint vector starts out uniformly random for motor commands toward each of the three locations on the screen, since a participant's initial movement biases are unpredictable. However, as the linguistic and contextual biases exert their influence on the visuomotor vector (as well as on one another), its activation pattern changes gradually and nonlinearly over time to conform to those biases. Since we treat evolving motor commands and evolving cognitive decisions as coextensive with one another (Cisek & Kalaska, 2005; Gold & Shadlen, 2000, 2001), we allow this visuomotor vector to send its biases to the integration vector just like all the other constraint vectors do. Therefore, an initially random motoric bias toward moving to the upper square can cooperate with a linguistic bias toward that same garden-path interpretation and result in a temporary “gang effect” whereby the simulated mouse trajectory curves upward before the disambiguating second PP vector eventually pulls everything its way. Following previous simulations of continuous motor movements (Spivey et al,. 2005; see also Godijn and Theeuwes’s, 2002, competitive integration model), it is directly from this visuomotor vector that we sampled a cascaded blend of the three motor commands. The distance (in x and y pixels) from the current simulated mouse location to each of the three potential destinations was calculated, and weighted by their corresponding activation values. The resulting [x,y] vector was scaled by one tenth of the activation of the most active visuomotor node to produce the direction and magnitude of the cooordinate transition for that time-step. Thus, an uncertain visuomotor vector, with near equal activations, would make a small movement on that time step, whereas a confident visuomotor vector, with only one substantially active node, would make a more sizeable movement on that time step.

Figure 3.6 shows the mean simulated computer-mouse movements from 100 runs of the model in each experimental condition. Only 50 time steps are plotted because some of the simulated trajectories took no more than 50 time steps. Closely matching the human data (Figure 3.1), the one-referent context simulation (Figure 3.6, upper panel) shows considerable divergence between the ambiguous- and unambiguous-sentence conditions, with the ambiguous-sentence resulting in a prolonged spatial attraction toward the incorrect destination. The two-referent context simulation (Figure 3.6, lower panel) shows no divergence between the ambiguous- and unambiguous-sentence conditions. These simulation results stand as an existence proof that a constraint-based model using dynamic competition between simultaneously partially-active representations (and continuous flow of those biases

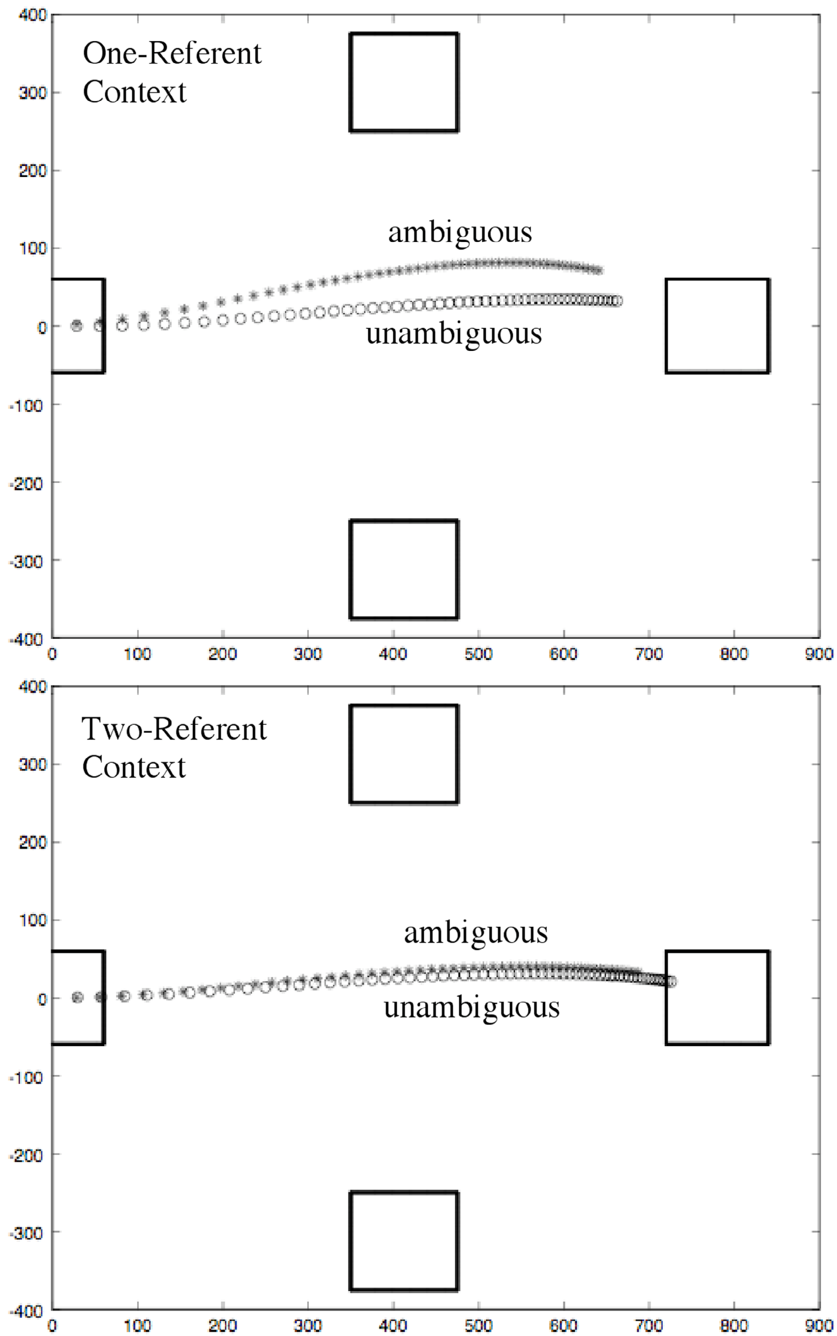

Figure 3.5. Constraint-based version of simulation. Trajectories are averaged from 100 runs of the integration-competition model, corresponding to each of the four cells in the Context x Ambiguity interaction. For the one-referent condition, as with the human data, substantial divergence between the ambiguous- and unambiguous-sentence conditions was observed (top panel). No substantial divergence occurred in the two-referent condition (bottom panel).

onto weighted-average motor commands) is able to account for the garden-path effect and its contextual modulation in this task.

However, since the unrestricted race model also combines multiple information sources (Traxler et al., 1998; van Gompel, 2001, 2005), it should also be able to account for the garden-path effect and its contextual modulation. In fact, the present simulation arrangement provides an opportunity to construct a version of the model that adheres to the claims of the unrestricted race account. Although the unrestricted race model of syntactic ambiguity resolution has yet to be explicitly implemented computationally for producing quantitative predictions, the present model architecture actually allows one to forgo the dynamic competition altogether and simply immediately select one or another alternative, depending on its activation—as posited by the unrestricted race account. Under these circumstances, only one of the visuomotor nodes will ever be allowed to drive the [x,y] coordinate changes during any phase of the spoken input. The resulting mouse-movements from such a simulation are mostly horizontal trajectories in the unambiguous conditions, and a combination of horizontal trajectories and quite angular trajectories in the ambiguous conditions, because movement will often be initially directed solely to the incorrect destination and then corrected to move toward the correct destination. When 100 simulations are averaged for each condition, as before, the results do a reasonable job of approximating the human data (Figure 3.7). (As before, only 50 time steps are plotted because some of the simulated trajectories took no more than 50 time steps.) Thus, it would appear that the unrestricted race account can produce an existence proof of about equal quality to that of the constraint-based model.

However, as noted in the introduction, the key differentiation between constraint-based models and the unrestricted race account is that the latter predicts a

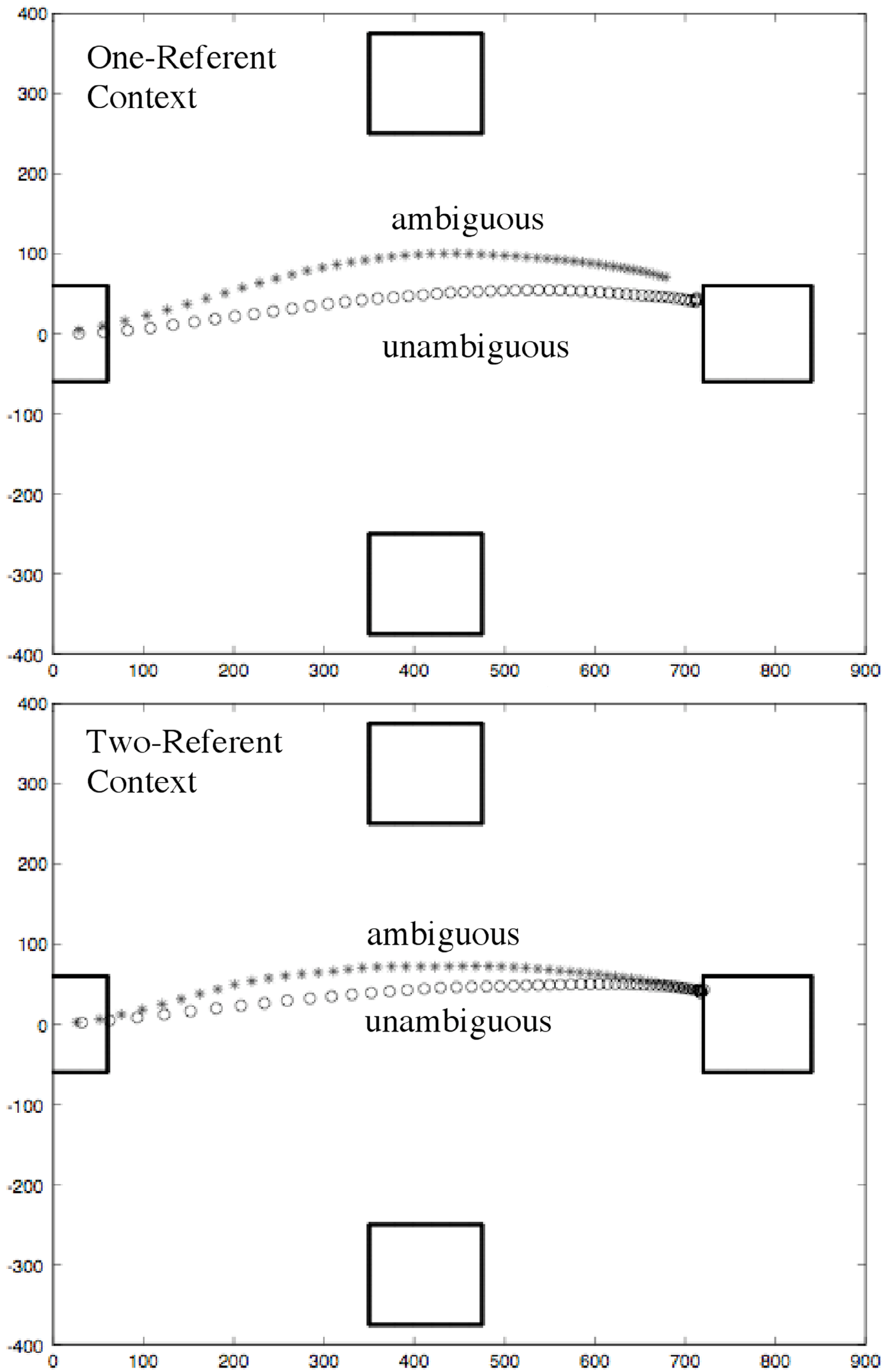

Figure3.7. Unrestricted Race version of the simulation. Trajectories are averaged from 100 runs of the integration-competition model in which only one visuomotor node was allowed to drive motor behavior at any one time.

bimodal distribution of garden-path magnitude because some trials involve a discrete commitment to an incorrect parse (thus requiring re-analysis) and others involve a discrete commitment to the correct parse from the beginning. As long as a significant garden-path effect is being observed across ambiguous and unambiguous conditions, then the distribution of garden-path magnitudes among the garden-path trials must be clearly shifted substantially away from that of the non-garden-path trials. Thus, if an ambiguous condition were composed of a combination of some of those garden-path trials and some of those non-garden-path trials, as predicted by the unrestricted race account, then it should exhibit a bimodal distribution of garden-path magnitude. Figure 3.8 examines this distribution by plotting the set of 100 simulated trajectories in the one-referent ambiguous-sentence condition. Since the unrestricted race simulation (lower panel) can only direct mouse movements toward one target location at a time (not toward weighted averages of target locations), it produces many perfectly horizontal trajectories and many quite angular trajectories that do not mimic well those of the human data (compare to Figure 3.3). The constraint-based simulation, by contrast, produces smoothly curved trajectories with a variety of shapes that closely resemble those of the human data.

When maximum deviation is calculated for each simulated trajectory in Figure 3.8, the distribution of curvature magnitudes is radically different for the two simulations. The constraint-based simulation produces a unimodal distribution of maximum deviation (Figure 3.9, upper panel) similar to that in the human data (Figure 3.4a), with kurtosis= 0.120, skewness= 0.030, and the bimodality coefficient, *b*= 0.311. In contrast, the unrestricted race simulation (Figure 3.9, lower panel) produces a clearly bimodal distribution of maximum deviation that does not fit the human data, with kurtosis= -0.995, skewness= -0.48, and the bimodality coefficient, *b*= 0.587. Thus, while the averaged-trajectory results (Figures 3.6 and 3.7) could perhaps not

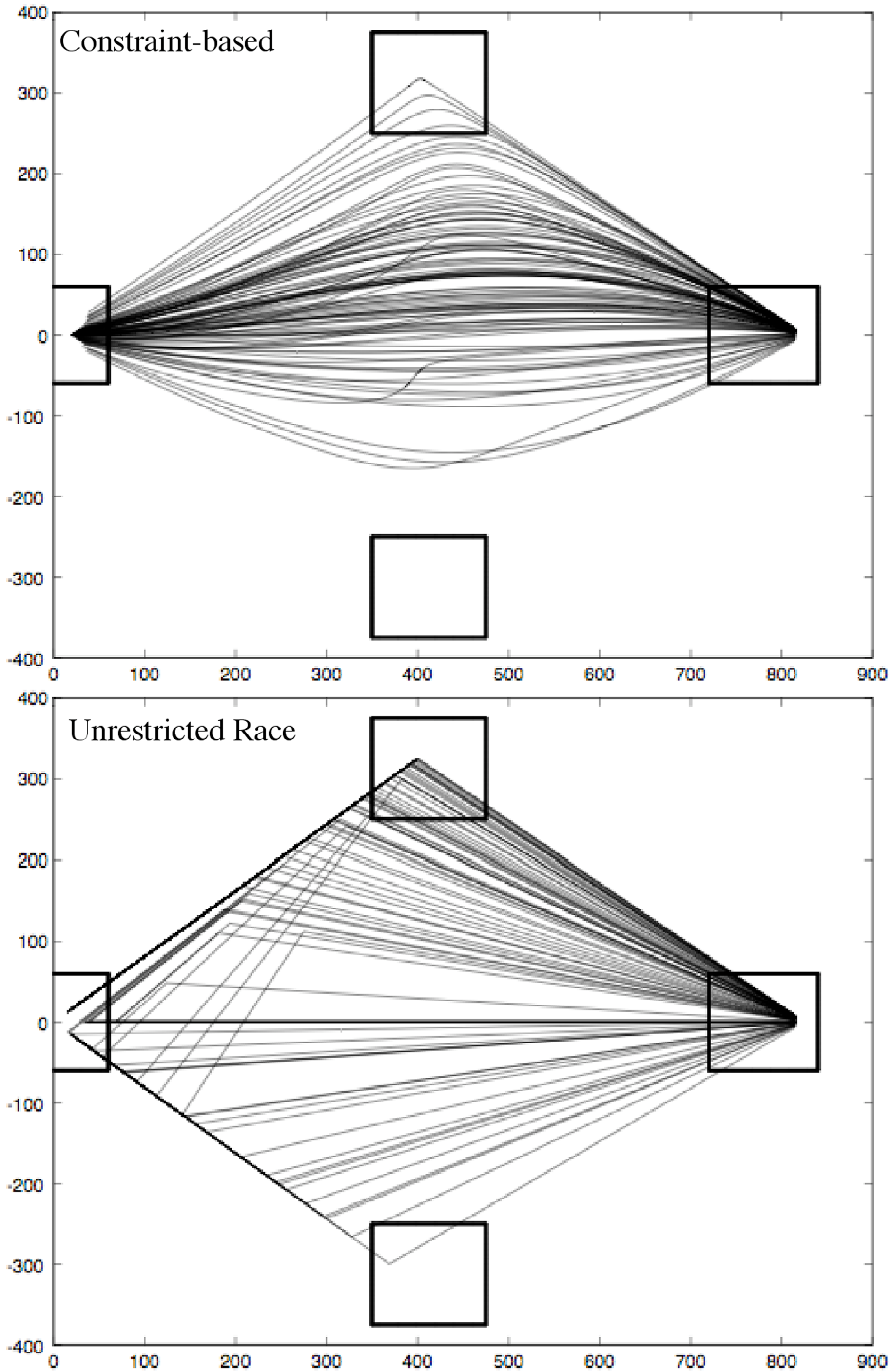

Figure 3.8. The trial-by-trial (n = 100) overlay of trajectories in the one-referent ambiguous-sentence (garden-path) condition, for the constraint-based and unrestricted-race versions of the simulation. Only the constraint-based simulation (upper panel) produces individual trajectories that resemble those of the human data.

unequivocally adjudicate between constraint-based and unrestricted race models, these distributional analyses show that the constraint-based model clearly outperforms the unrestricted race model in simulating the temporal dynamics of computer-mouse movements during spoken language comprehension in a visual context. It is possible that additional parameters could be added to the unrestricted race simulation to smooth out its individual trajectories, and perhaps even reduce the bimodality in its distribution of curvature magnitudes. For that theoretical position to maintain its viability in the face of present findings, explicit quantitative simulations like these, with fits to human data, will be necessary to demonstrate how such modifications

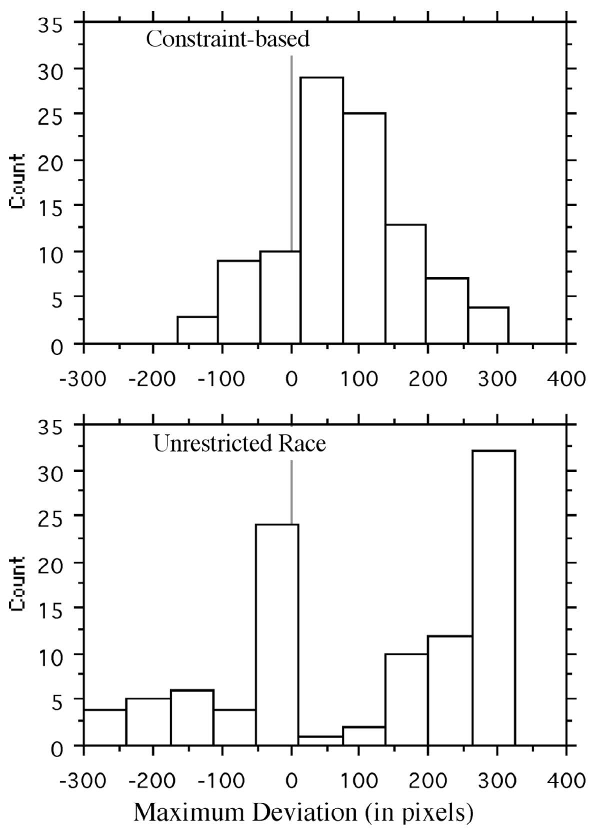

Figure 3.9. The distributions of simulated trajectory curvature magnitudes from Figure 3.8, calculated as maximum deviation (in y-pixels) from a straight line. The constraint-based simulation produces a unimodal distribution much like that in the human data (Figure 3.4A), whereas the unrestricted race simulation produces a clearly bimodal distribution of many flat non-garden-path trajectories and many upwardly-angled garden-path trajectories.

could be successful.

Experiment 3

The distribution of trajectory-curvature elicited in the garden-path condition in Study 1 provides support for gradation in the magnitude of the garden-path effect and is consistent with graded competition between two simultaneously active representations. Moreover, Study 2 illustrates that a computational implementation of the constraint-based account is capable of producing smoothly curved trajectories mirroring those found in Study 1. The combined results of these two studies provide support for a model of syntactic processing whereby multiple representations of an ambiguity are simultaneously active and compete for activation across time based on the information available to the system. Importantly, the distribution of maximum deviation values in Study 1 (an index of garden-path magnitude) is unimodal and is thus difficult to reconcile with the unrestricted race model which predicts a bimodal distribution of garden-path magnitudes that would correspond to one population of trials where participants were garden-pathed and a separate population of trials where they were not.

Although the syntactic ambiguity manipulation in which we are primarily interested did not appear to produce a bimodal distribution of garden-path magnitudes, it is possible that the mouse-tracking paradigm and/or the bimodality coefficient are not sensitive enough to illuminate an underlying bimodal distribution of responses to garden-path sentences. That is, perhaps factors that are unrelated to language, such as the kinematics of wrist and/or hand movements along the horizontal movement plane, may be limiting the motor output in a way that prevents a bimodal distribution of trajectory curvatures from emerging. Additionally, assessing the number of modes within a distribution is difficult given the current statistical techniques available, and one concern is that the methods we used to assess the number of modes in the garden-path distribution (all of which support the conclusion that the distribution of garden-path magnitudes is unimodal) were not sensitive enough to detect bimodality in the distribution should it actually exist. In order to allay such concerns, we created a purely visuomotor experimental task with conditions that should produce mouse movements that are consistent with the various parsing models discussed above. Should the conditions employed here actually produce such mouse movements, it will then be possible to compare their distributions to the previously observed distributions of mouse movements in order to determine which parsing model best characterizes the distribution of responses to garden-path sentences.

In this control study, participants were presented with a scene consisting of four squares corresponding to each of the four possible object locations in Study 1. After being instructed to "Click on the green square," participants clicked on the square located in the center of the left edge of the display to begin a trial. They were subsequently presented with another green square to move to and click, in one of the three remaining object locations. For "garden-path" trials, the target green square appeared at the top-center of the display, and red squares appeared at the right- and bottom-centers of the screen. However, once the participant's mouse moved outside of the start box in pursuit of a click on that upper green square, the green square changed to a red square and the red square located at the right-center of the display changed to a green square. This switch required the participant to alter an initial up-rightward diagonal movement toward the top-center of the screen (just like toward the incorrect destination in Study 1) to a rightward (and somewhat downward) movement directed at the green square in its new location (the same location as the correct destination in Study 1). This color-switch thus simulated a situation whereby one discrete representation was initially active, issuing a motor command to move to the upper square, followed by a separate discrete representation issuing a command to move to the right square. A baseline "no-switch" condition was included in which the green square appeared at the right of the display, with red squares appearing in the other two locations and no switch ever occurring. This condition mirrored the unambiguous-sentence conditions in Study 1, requiring a simple left-to-right movement with no activation of any analysis corresponding to the incorrect destination. In a third condition, participants were presented with a set of "competition" trials in which a red square appeared at the bottom-center location, a green square at the right-center location, and a greenish-blue square at the top-center location. This condition corresponds to the constraint-based prediction that multiple syntactic representations may be partially active at the same time (McRae et al., 1998), and issuing multiple motor commands at the same time (Cisek & Kalaska, 2005), which result in a continuously updated movement vector that is an average of the multiple motor commands, dynamically weighted by the changing activations of their corresponding linguistic representations.

The distribution of garden-path (switch) trials combined with baseline (no-switch) trials should produce the response distribution that the unrestricted race account predicts for syntactically ambiguous sentences—one in which a garden-path would either occur due to the discrete selection of the ultimately incorrect representation, or would not occur, due to the discrete selection of the ultimately correct alternative. By examining the distributional properties of the maximum deviation values produced by the garden-path and non-garden-path trials, together, we can thus determine whether or not the statistical techniques we used to assess the bimodality of the garden-path distribution in Study 1 are capable of detecting bimodality in a case where the response distribution should clearly be bimodal (as it should be if the unrestricted race account were accurate). Moreover, because we also included a condition that should induce graded-competition similar in nature to the competition between syntactic alternatives posited in Study 1, it is possible to compare the properties of the garden-path distribution in Study 1 to the properties of both the unrestricted race distribution and the competition distributions in this study to determine which distribution best characterizes the garden-path distribution created by the presence of syntactic ambiguity.

Predictions

In relation to the actual shapes of the trajectories, it was predicted first that movement in the baseline no-switch condition would mirror the average rightward horizontal movement produced by the unambiguous-sentence conditions in Study 1, and that significant y-coordinate divergence would be seen between the baseline (no-switch) trials and both the competition and garden-path (switch) trials. It was also predicted that the presence of a color-switch would induce a strong garden-path effect not unlike the most extreme garden-path trials in the garden-path condition from Study 1 (most evident on Figure 3.3). In relation to the distribution of maximum deviation values, it was predicted that the combined switch and no-switch distribution would be bimodal, as indexed by the bimodality coefficient (*b* should be > .555), whereas the competition distribution would be unimodal. Most importantly, based on the results of Studies 1 and 2, we also predicted that the garden-path distribution from Study 1 would have properties roughly identical to the competition condition in this present study and that the shapes of the two distributions would be indistinguishable as determined by the Kolmogorov-Smirnov test. Such a result would suggest that the garden-path trajectories in Study 1 arise from a highly-active signal to move rightward (such as a green target square in that location) *accompanied simultaneously* by a partially-active signal to move upward (such as a greenish blue square in that location).

Method

Participants

A separate group of 26 right-handed Cornell undergraduates (*M* = 19.6 years, *SD* = 1.1) participated in this study for extra course credit.

Materials

As in Study 1, all stimuli were presented using Macromedia Director MX (display resolution = 1024 x 768), and mouse movements were recorded at an average sampling rate of 40 Hz. All 24 experimental trials and 48 filler trials involved the presentation of three 1.5 X 1.5 inch squares that were either red, green, or greenish-blue, depending on the experimental condition or type of filler trial. For all trials, experimental and filler alike, one square always appeared at the top-center of the display, another at the center of the right side of the display, and another at the bottom-center. For all trials, a 1.5 X 1.5 inch start-box appeared at the center of the left side of the screen and contained the words “Click Here to Begin.” Each square subtended an average of 4.64 degrees in width by 4.64 degrees in height of visual angle. The start-box, located on the far left of the screen, subtended 12.69 degrees of visual angle from the center of the screen, the square on the far right of the screen subtended 12.69 degrees of visual angle from the center of the screen, and the squares in the bottom- and top-center positions each subtended 10.37 degrees of visual angle from the center of the screen. Figure 3.10 illustrates the relative locations of the squares.

On all trials, the green square was always the target square on which participants were to click. For 24 of the filler trials, the green square appeared at the top-center of the screen, with red squares occupying the right- and bottom-center square locations, and for the other 24 filler trials, the green square appeared at the bottom-center of the display, with red squares occupying the right- and top-center square locations. Given that the goal of this study was to examine mouse-movements that are analogous in nature to the movements produced during the processing of the experimental sentences in Study 1, on the 24 experimental items, the green square appeared initially, or ended-up, at the right-center location of the display—a location that corresponds to the location of the correct destination in Study 1. As in Study 1, then, the movement always initiated at the left-center of the display (at the start-box in this study or at the target referent-object in Study 1) and progressed rightward, terminating at the right-center of the display.

Of the 24 experimental items, eight were garden-path (switch) trials, eight were baseline (no switch) trials, and the remaining eight were competition trials. The characteristics of these trial-types are described in the introduction to this study.

Procedure

Participants were instructed simply to "click on the green square with as much accuracy as possible." At the beginning of each trial the start-box appeared, and upon clicking on it, three 1.5 X 1.5 inch white boxes, demarcated by black dotted lines on all four sides (as in Figure 3.10), subsequently appeared. Participants were thus informed of exactly where the colored squares were about to appear, but were prevented from planning any course of action because they did not know the location of the green square. After one second, the empty boxes were replaced with three colored squares corresponding to one of the trial-types listed above. The cursor was frozen at the exact location in the start-box where the trial-initiating click occurred, and remained frozen throughout the one-second delay up until the colored squares appeared, at which point the participant was able to move the cursor with the mouse to click on the green square. This delay of movement functioned to prevent any anticipatory movement. Once the participant clicked on the green square, the current display disappeared and participants were presented with the start-box for the next trial. The 72 items were presented in one experimental block, and the order of item presentation was randomized per participant.

Results and Discussion

Trajectory Analyses

Mouse movements were recorded throughout the duration of the trial, beginning with the point at which the colored squares appeared up until the point at which the green square was clicked. The x,y coordinates from the trajectories of each experimental trial were graphed individually in order to identify any aberrant movements or any trials on which participants clicked one of the incorrect (non-green) squares before ultimately clicking on the green square. No trial contained an aberrant or nonsensical movement, and only 1.4% of the experimental trials were excluded from the analyses presented below because they involved a mouse-click of a non-green square. Each analyzable trajectory was time-normalized to 101 time-steps as in Study 1.

Figure 3.10 displays the averaged time-normalized trajectories produced by each experimental condition. The average movement produced by the baseline no-switch condition corresponds to the average movement on the unambiguous-sentence conditions in Study 1, with a rather straight horizontal movement from left to right. It also appears that the switch condition produced a large number of extreme garden-paths, with trajectories initially traveling toward the upper square (corresponding to the incorrect destination in Study 1) before being redirected to the ultimately correct destination after the color-switch occurred. Moreover, it appears that the competition condition produced, on average, a more subtle graded curvature toward the location of the greenish-blue square, as is the case for the garden-path condition in Study 1 (compare to Figure 3.1, upper panel).

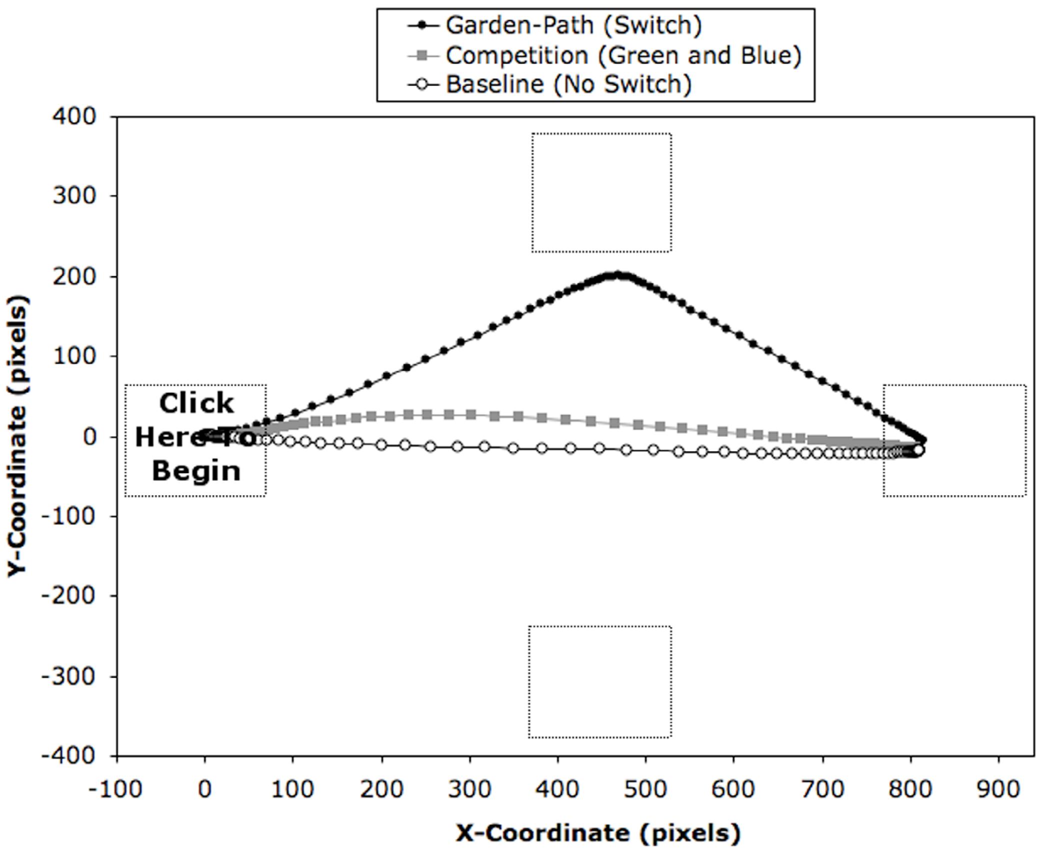

Figure 3.10. Visuomotor Control study. The mean mouse-movement trajectory for the “Garden-path” condition shows a sharply-angled curvature, while the “Competition” condition shows subtle graded curvature, and the “Baseline” condition shows a genuinely flat trajectory.

In order to assess the degree to which the x,y coordinates of the trajectories produced in the garden-path and competition conditions diverged significantly from the baseline (no-switch) condition, we conducted a *t*-test at each of the 101 time-steps for the x- and y-coordinates, separately. As in Study 1, an observed divergence was not considered significant unless the coordinates between the baseline condition and the competition or garden-path conditions elicited p-values < .05 for at least eight consecutive time-steps. Also as in Study 1, the x-coordinates of the elicited trajectories are solely indicative of velocity toward the correct destination, and the y-coordinates are solely indicative of spatial attraction toward the ultimately incorrect destination.

Average trajectories in the garden-path condition traveled rightward significantly faster than they did in the baseline no-switch condition from time-steps 18-45, as indexed by the x-coordinate comparisons, all t's > 2.14, all p's < .05, average d = 1.43. However, from time-steps 50-85, this difference reversed, with the average trajectory in the baseline condition traveling rightward toward the location of the ultimately correct destination more quickly than did the average trajectory in the garden-path condition, all t's > 2,07, all p's < .05, average d = 1.67. As is evident in Figure 3.10, there was also significantly more spatial attraction toward the top of the screen, corresponding to the location of the incorrect destination in Study 1, in the garden-path condition than there was in the baseline condition, as indexed by the significant y-coordinate comparisons from time-steps 19-101, all t's > 2.57, all p's < .05, average d = 2.17. When considering the x- and y-coordinate analyses together, then, it appears that for the first half of the average trial in the garden-path condition, participants made a quick movement upward toward the initial location of the green square, but after the switch occurred, the movement was redirected toward the new location of the green square at the right-center of the display. On average, these re-directed trajectories arrived at the ultimately correct destination later than the average baseline no-switch trajectories. These results support the presence of a clear garden-path effect in the switch condition, consisting of one discrete motor command to move the cursor to the upper square, quickly replaced by a different motor command to move the cursor to the rightmost square.

Substantial x-coordinate divergence also occurred between the average trajectories in the baseline condition and in the competition condition from time-steps 47 to 82, all t's > 2.10, all p's < .05, average d = .451, with trajectories in the baseline condition being closer to the correct destination (green square) than the competition condition trajectories at each of those time-steps. Substantial y-coordinate divergence was observed as well between the average trajectories in the baseline condition and the competition condition trajectories from time-steps 28-78, all t's > 2.16, all p's < .05, average d = .744. At each of those time-steps, the trajectories in the competition condition were significantly closer to the location of the competing greenish-blue square at the top of the screen than were trajectories in the baseline condition.

Comparisons between the x-coordinates of the competition and garden-path conditions revealed that from time-steps 20-49, all t's > 2.08, all p's < .05, average d = 1.272, the garden-path condition elicited trajectories that traveled rightward more quickly than did the competition condition. As was the case with the x-coordinate comparisons between the averaged baseline and garden-path condition trajectories, however, from time-steps 53-81, the trajectories in the garden-path condition traveled rightward toward the ultimately correct destination more slowly than did the average trajectories in the competition condition, all t's > 2.32, all p's < .05, average d = 1.264. Additionally, the y-coordinate comparisons revealed that the average trajectories in the garden-path condition were significantly closer to the top of the screen than were the average trajectories in the competition condition for time-steps 19-101, all t's > 2.09, all p's < .05, average d = 1.779.

The results of these analyses confirm that the mouse-tracking paradigm is capable of producing average movements that correspond to the predictions of various models of syntactic processing. First, the average movement in the baseline (no-switch) condition is one characterized by a rightward movement traversing the horizontal movement plane between the start- and end-points. Such a movement is consistent with what was observed for the unambiguous-sentence conditions in Study 1, in which the absence of a syntactic ambiguity prevented any noteworthy spatial attraction toward any other location in the display. This movement pattern also corresponds to the portion of syntactically ambiguous trials that the unrestricted race model predicts would not need reanalysis because the ultimately correct representation of a sentence's syntactic structure was initially discretely selected. Secondly, the garden-path condition produced trajectories that moved up and rightward more quickly than either of the other two conditions for the first half of the trial, but not for the second half. Such a trend suggests that, relative to the competition condition, participants were initially strongly biased toward the top-center of the screen and moved accordingly, but that upon processing the color switch, redirected their movement toward the new location of the green square. This pattern is consistent with the portion of syntactically ambiguous trials that the unrestricted race model predicts would need reanalysis because the incorrect representation of a sentence's syntactic structure was initially selected.

Most interesting, however, is that the average movement in the garden-path condition of Study 1 (Figure 3.1, upper panel) does not resemble the average movement in either the garden-path switch or the baseline no-switch conditions here. Instead, the average movement in the presence of a syntactically ambiguous sentence in Study 1 looks a lot like the competition condition in this study, where a more subtle upward curvature toward the competing location occurs, and where the velocity of the trajectory toward the ultimately correct location is slower when competition occurs than when the sentence is unambiguous (thus involving little or no competition). Of course, these across-study comparisons are qualitative in nature, but given that the average movements across each of the three conditions in this study map onto the predictions of various parsing models, cross-study quantitative comparisons of the distributions of mouse movements in the various conditions are warranted.

Distributional Analyses

Figure 3.11 displays the distributions of curvature magnitudes for each of the three conditions employed here, along with a fourth distribution that was created by combining the distributions of the baseline no-switch and garden-path switch trials. The descriptive statistics associated with each of the four distributions can be found in Table 3.2. The distribution produced by the baseline no-switch condition is unimodal, bimodality coefficient $b < .555$, and appears to vary normally around the mean. By contrast, the garden-path trial distribution elicited a $b > .555$. This distribution, alone, has a shape that would be predicted by the unrestricted race model in the presence of a syntactic ambiguity with constraints that relatively strongly bias the system toward the ultimately incorrect syntactic alternative. The large rightmost mode corresponds to a majority of trials on which reanalysis is needed due to a garden-path, and the much smaller leftmost mode corresponds to the proportion of trials in which participants discretely selected the ultimately correct alternative from the beginning. That is, when participants are faced with a situation that corresponds to a true garden-path where basically only one representation can be discretely considered at a time, distributions of mouse-movements are still bimodal. Given the disparity in the descriptive properties between the distributions of garden-path-strength values in the one-referent ambiguous-sentence condition in Study 1 and the garden-path switch distribution here, however, it is safe to say that the distribution that would be predicted by the unrestricted race account is an inappropriate characterization of the language-related garden-path phenomenon. Furthermore, the Kolmogorov-Smirnov test demonstrates that the shapes of those two distributions are significantly different, $p < .0005$.

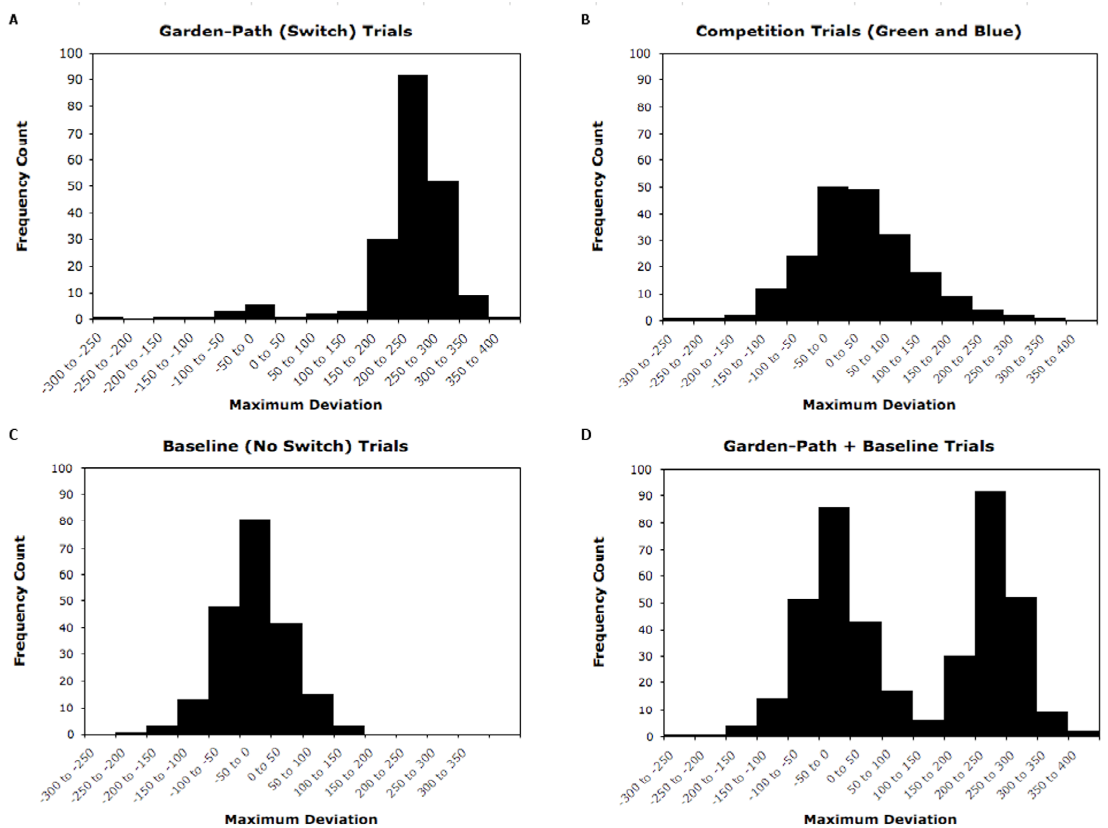

Figure 3.11. In the visuomotor control study, the distributions of maximum deviation (in y-pixels) from a straight line reveal bimodal distributions when baseline non-garden-path trials are combined with garden-path trials (Panel D), as well as in the garden-path condition alone (Panel A). In the competition condition alone, corresponding to a constraint-based account of multiple representations being partially active at the same time, the distribution is unimodal with a mean slightly greater than zero. In the baseline condition alone (Panel C), corresponding to an unambiguous-sentence condition, the distribution is unimodal with a mean of zero.

We combined the baseline and garden-path switch trials into one distribution in order to approximate a distribution comprising some trials that required a re-analysis and some trials that did not require a re-analysis. The unrestricted race model would predict such a distribution in the presence of an ambiguity with constraints that support each of the alternatives roughly equally. The combined distribution is clearly bimodal, as evident from visual inspection of Figure 3.11d, and by the fact that it elicited a b-value > .555 (Table 3.2). As was the case with the distribution of garden-path magnitudes in the switch condition alone, the properties of this combined distribution also do not align well with the properties of the distribution produced by the presence of syntactic ambiguity in Study 1. Moreover, a Kolmogorov-Smirnov test found that the shape of the Study 1 garden-path distribution (Figure 3.4a) is significantly different from the shape of the Study 3 combined distribution (Figure 3.11d), $p < .0005$.

Table 3.2

Maximum deviation statistics for the four distributions of trials in Study 3.

Condition	n	Mean	*SD*	Skewness	Kurtosis	Bimodality (*b*)
Baseline (No Switch)	206	-24.43	56.93	-.01	.71	.267
Garden-path (Switch)	202	214.35	85.99	-2.72	9.27	.680
Competition (Bluish-green)	205	19.96	90.88	.39	.90	.292
Combination (GP + Baseline)	408	94.18	140.40	-.02	-1.43	.630

In contrast, the properties of the Study 3 competition distribution (Figure 3.11b) closely mirror those of the distribution of curvature magnitudes in the garden-path condition of Study 1 (Figure 3.4a). As was the case with the garden-path distribution in Study 1, bimodality was not detected in the competition condition here, $b < .555$. The means of the garden-path distribution in Study 1 ($M = 23.82$) and the competition distribution here (*M* = 19.96), along with the standard deviations (Study 1 *SD* = 91.87, Study 3 *SD* = 90.88), are almost identical, and the shapes of the distributions are statistically indistinguishable by the Kolmogorov-Smirnov test, $p > .1$. Unlike the distributions that are predicted by the unrestricted race model, then, the distribution that actually does accommodate garden-path effects on syntactically ambiguous sentences is one produced by continuously graded competition.

The results of this study demonstrate that the mouse-tracking technique employed here can produce average movements that correspond to what would be predicted by various parsing models. The presence of bimodal distributions of movements in the garden-path switch and combination distributions demonstrates not only that the mouse-tracking technique can produce a bimodal distribution when one is expected, but also that the statistics we employed to assess the number of modes within the various distributions are sensitive enough to detect bimodality when it is present. Therefore, the conspicuous absence of evidence for bimodality in the distribution of garden-path magnitudes in Study 1 (Figure 3.4a) is likely due to there not being any bimodality in the way that garden-path sentences are processed.

General Discussion

We have presented converging evidence from computer-mouse movements and model simulations in the visual-world paradigm that provide evidence in favor of the constraint-based account of sentence processing, in which multiple partially-active syntactic alternatives compete with one another via support from a variety of information sources (e.g., Elman et al., 2004; MacDonald et al., 1994; Spivey & Tanenhaus, 1998; Trueswell et al., 1994). Interestingly, it is a *visual* context that is influencing this linguistic competition process. The real-time information flow between visual and linguistic information is particularly well illustrated in the continuous trajectories of computer-mouse movements in Study 1. Although these movements of the hand are initiated slightly later than the first eye movement of a scan path, they are considerably smoother and less ballistic than saccades, and they typically reveal their graded spatial attraction around the same period of time (relative to the spoken sentence) that many of the critical fixations of competing objects tend to occur in the visual world paradigm (Chambers et al., 2004; Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999). Our results clearly demonstrate that, while a participant drags an object toward its syntactically-correct destination, variation of the visual context modulates the tendency to move that object partly in the direction of its garden-path destination as well (Figures 3.1 and 3.2).

According to constraint-based accounts of sentence processing, part of why visual context is able to immediately influence the resolution process is precisely because the correct alternative was never summarily discarded during the comprehension system's partial foray down the garden path. In fact, van Gompel himself reports evidence that activation of the *inappropriate* parse of a temporary syntactic ambiguity lingers for long enough after the sentence to exert syntactic priming on the production of a subsequent sentence (van Gompel, Pickering, Pearson, & Jacob, 2006). If statistical, semantic, and structural biases were to persuade the processing system to *eliminate* the syntactic alternative that would have turned out to be the correct one, then new information that supported that now-absent alternative would have no available representation to receive said support. However, if those other constraints merely *inhibited* the graded activation of that alternative, then a strongly supportive new constraint could perhaps be influential enough to bring that suppressed (but not eliminated) alternative back to a prominent activation level. For example, look at what the integration-competition model does when it displays a substantial garden-path effect. Initially, the incorrect alternative prevails over the correct alternative by a substantial margin of activation. During that portion of the sentence, there is little competition and hence little processing difficulty, as the whole system settles on the wrong parse. However, when disambiguating information, such as the second PP in example 2a, provides evidence discretely in favor of the correct alternative, the overturning of that incorrect activation pattern involves a laborious and time-consuming competition process (see Green & Mitchell, 2006). Now compare that to what the model does when something like visual context prevents a substantial garden-path. Some local constraints support the wrong syntactic alternative and the contextual constraint supports the correct one, so the early portion of the sentence now exhibits a moderate amount of competition and processing difficulty, which does *not* get fully resolved before later portions of the sentence are heard or read. As a result, the model has not dug in its heels to defend either alternative, and when the disambiguating portion of the sentence is encountered, those new biases can now take over relatively smoothly.

This gradiency in the simultaneous activation of syntactic alternatives may provide a turning point in the debate between re-analysis theories (whether of the syntax-first variety or the unrestricted race variety) and competition theories (typically, constraint-based). Previous work has shown that in a one-referent visual context, it is about half of the time that participants fixate that incorrect destination, and the other half of the time they look only at the correct destination (Spivey et al., 2002; Tanenhaus et al., 1995). Therefore, this would appear to be the ideal case where the two syntactic alternatives are near 50/50 in their salience. Hence, the unrestricted race theory should predict a clearly bimodal distribution, with about half of the trials exhibiting dramatic spatial attraction to the incorrect destination, and the other half showing none at all. In fact, that pattern is exactly what the eye-tracking data show, because saccadic eye movements tend to be quite ballistic. However, in Study 1, our distribution of continuous computer-mouse trajectories in the one-referent ambiguous-sentence condition showed no evidence of such a bimodal distribution (Figure 3.3). The degree of curvature among the trajectories was distributed in a clearly unimodal fashion (Figure 3.4a). That is, the graded spatial attraction effects elicited in this condition came not from two different types of trials (some engaging a re-analysis mechanism and some not doing so) but from a single population of trials (all engaging the same competition process). Moreover, the shape of the distribution was not significantly different from that of a distribution of curvature values elicited by control sentences (a condition where no theory would predict a bimodal distribution).

Furthermore, in Study 2, an integration-competition simulation of continuously emitted [x,y] coordinate changes (inspired by constraint-based accounts of sentence processing), using a weighted combination of the active alternatives to produce a blend of movements, provided a close fit to the actual mouse-movement trajectories (compare Figures 3.3 and 3.8a). Much like the distribution of curvature magnitudes from the human data (Figure 3.4a), these simulated trajectories exhibited curvature magnitudes that formed a unimodal distribution (Figure 3.9a). In contrast, when the same model used only one syntactic alternative at a time to drive [x,y] movements, corresponding to the unrestricted race theory, the pattern of simulated trajectories did not match well with the human data (Figure 3.8b), and its curvature magnitudes formed a clearly bimodal distribution (Figure 3.9b).

Finally, in Study 3, a visuomotor control task demonstrated that when a signal is initially misleading about where to move the mouse, our mouse-tracking paradigm is able to reveal the dramatically curved trajectories that result. And when a distribution of curvature magnitudes contains some of those dramatically curved trajectories and also some very straight trajectories, our tests for bimodality can detect the presence of two separate populations of trials. Therefore, the fact that we did not find evidence for bimodality in Study 1 is indeed informative. It suggests that the garden-path trajectories in the one-referent ambiguous-sentence condition come from a single population of sentence processing events. Also, the competition condition in Study 3, where the upper distractor square was similar in color to the target square, exhibited a unimodal distribution of graded curvature magnitudes (Figure 3.11b) that was not statistically different from the distribution of curvature magnitudes in the garden-path condition of Study 1 (Figure 3.4a). This observation lends further support to conceiving of these garden-path effects as resulting from competition between simultaneously partially-active representations.

Overall, the results described here tie in nicely with converging evidence for a close-knit relationship between language processing, visual perception, and motor action (e.g., Barsalou, 1999; Chambers et al., 2004; Glenberg & Kaschak, 2002; Pulvermüller, 1999; Spivey et al., 2005; Zwaan & Taylor, 2006). And if perceptual and motor processes rely on distributed graded activations of multiple representations evolving in real-time (e.g., Paninski et al., 2004; Rolls & Tovee, 1995), perhaps it should not be surprising that tightly-yoked linguistic processes would follow suit. This, of course, would not be the first time that cognitive psychology has witnessed the gradual blurring of a historical dichotomy between two categorically different perceptual processes (e.g., a garden-path event and a non-garden-path event). For example, in the 1970's, necessary and sufficient conditions for discretely defining an exemplar as either a member or a non-member of a category gave way to the concept of prototype-based *graded membership* in a category (Rosch, 1973; Zadeh, 1975). And in the 1990's, the categorical distinction between parallel and serial visual search gave way to a *continuum* of search efficiency (Duncan & Humphreys, 1989; Wolfe, 1998).

In fact, hints of this kind of gradiency have been showing up in a number of recent approaches to syntax, in the form of either probabilistic or underspecified representations (e.g., Bod, Hay, & Jannedy, 2003; Ferreira, Bailey, & Ferraro, 2002; Hale, 2006; Jurafsky, 1996; Levy, 2006; Weinberg, 1993; see also Tabor, Galantucci, & Richardson, 2004). This departure from traditional frameworks (which required a discrete commitment to a determinate parse), and the present gravitation toward continuous dynamical frameworks for syntax (Culicover & Nowak, 2003; Tabor & Hutchins, 2004), are bringing the field of sentence processing in line with the growing successes of dynamical-systems accounts of cognition in general (e.g., Port & Van Gelder, 1995; Spivey, 2007; Ward, 2002). As a result, the theoretical treatment of syntactic garden-path effects is likely to require something of a reformulation. Rather than conceiving of the pursuit of a syntactic structure as an all-or-nothing process, on which a discrete re-analysis either will or will not be required (Traxler et al., 1998; van Gompel et al., 2001, 2005), the results we report point to a gradiency in the degree to which an incorrect syntactic structure is pursued in conjunction with the correct syntactic structure, which is consistent with competition-based accounts of constraint-based sentence processing (e.g., Elman et al., 2004; Green & Mitchell, 2006; MacDonald et al., 1994; McRae et al., 1998; Tabor & Tanenhaus, 1999).

REFERENCES

Abrams, R. A., & Balota, D. A. (1991). Mental chronometry: Beyond reaction time. *Psychological Science, 2*, 153-157.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419-439.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264.

Altmann, G. T. M., & Steedman, M. J. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191-238.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3-73). Cambridge, UK: Cambridge University Press.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the growth of cognition* (pp. 279-362). New York, NY: Wiley.

Bod, R., Hay, J. & Jannedy, S. (Eds.). (2003). *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory and Language, 51*, 495-522.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language, 33*, 251-283.

Britt, M. A., Perfetti, C. A., Garrod, S., & Rayner, K. (1992). Parsing in discourse:

Context effects and their limits. *Journal of Memory and Language, 31*, 293-314.

Budiu, R., & Anderson, J. R. (2004). Interpretation-based processing: A unified theory of semantic sentence comprehension. *Cognitive Science, 28*, 1-44.

Chambers, C., Tanenhaus, M., & Magnuson, J. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*, 687-696.

Cisek, P., & Kalaska, J. (2005). Neural correlates of reaching decisions in dorsal premotor cortex. *Neuron, 45*, 801–814.

Clifton, C., Traxler, M., Mohamed, M., Williams, R., Morris, R., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 49*, 317-334.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Culicover, P., & Nowak, A. (2003). *Dynamical grammar*. New York, NY: Oxford University Press.

Dale, R., Kehoe, C. E., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing exemplars. *Memory and Cognition, 35*, 15-28.

Darlington, R. B. (1970). Is kurtosis really "peakedness"? *American Statistician, 24*, 19-22.

DeCarlo, L. T. (1997). On the meaning of kurtosis and use of kurtosis. *Psychological Methods*, *2*, 292-307.

Desmet, T., De Baecke, C., & Brysbaert, M. (2002). The influence of referential discourse context on modifier attachment in Dutch. *Memory and Cognition, 30*, 150-157.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention.

Annual Review of Neuroscience, 18, 193-222.

Dietrich, E., & Markman, A. B. (2003). Discrete thoughts: Why cognition must use discrete representations. *Mind and Language*, *18*, 95-119.

Doyle, M., & Walker, R. (2001). Curved saccade trajectories: Voluntary and reflexive saccades curve away from irrelevant distractors. *Experimental Brain Research, 139*, 333-344.

Duncan, J., & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433-458.

Elman, J., Hare, M., & McRae, K. (2004). Cues, constraints, and competition in sentence processing. In M. Tomasello & D. Slobin (Eds.), *Beyond nature-nurture: Essays in honor of Elizabeth Bates* (pp. 111-138). Mahwah, NJ: Lawrence Erlbaum Associates.

Farmer, T. A., Cargill, S. A., Hindy, N. C., Dale, R., & Spivey, M. J. (in press). Tracking the continuity of language comprehension: Computer-mouse trajectories suggest parallel syntactic processing. *Cognitive Science*.

Ferreira, F., Bailey, K. G. D, & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11-15.

Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 248-368.

Fodor, J. D., & Ferreira, F. (Eds.) (1998). *Reanalysis in sentence processing*. Netherlands: Kluwer.

Frazier, L. (1998). Getting there (Slowly). *Journal of Psycholinguisic Research, 27*, 123-146.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movement in the analysis of structurally ambiguous-sentences. *Cognitive Psychology, 14*, 178-210.

Gaskell, M. G., & Marslen-Wilson, W. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science, 23*, 439-462.

Glenberg, A., & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin and Review*, *9*, 558-565.

Godijn, R., & Theeuwes, J. (2002). Programming of endogenous and exogenous saccades: Evidence for a competitive integration model. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 1039-1054.

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature, 404*, 390-394.

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences, 5*, 10-16.

Goodale, M. A., Pelisson, D., & Prablanc, C. (1986). Large adjustment in visually guided reaching do not depending on vision of the hand or perception of the target displacement. *Nature, 320*, 748-750.

Green, M., & Mitchell, D. (2006). Absence of real evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language, 55*, 1-17.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science, 30*, 643-672.

Hare, M., Tanenhaus, M. K., & McRae, K. (2007). Understanding and producing the reduced relative construction: Evidence from ratings, editing and corpora. *Journal of Memory and Language, 56*, 410-435.

Hindle, D., & Rooth, M. (1993). Structural ambiguity and lexical relations. *Computational Linguistics, 19*, 103-120.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*, 137-194.

Knoeferle, P. & Crocker, M. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science, 30*, 481-529.

Levy, R. (in press). *Expectation-based syntactic comprehension*. Manuscript submitted for publication. *Cognition.*

Luce, P., Goldinger, S., Auer, E., & Vitevitch, M. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics*, *62*, 615-625.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676-703.

McClelland, J. L., & Elman, J. L. (1986). Interactive processes in speech perception: The TRACE model. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing, VOL II* (pp. 58-121). Cambridge, UK: Cambridge University Press.

McRae, K., Spivey-Knowlton, M., & Tanenhaus, M. (1998). Modeling the effects of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 37*, 283-312.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Paninski, L., Fellows, M., Hatsopoulos, N., & Donoghue, J. (2004). Satiotemporal tuning of motor cortical neurons for hand position and velocity. *Journal of Neurophysiology*, *91*, 515-532.

Port, R., & van Gelder, T. (Eds.). (1995). *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*, 253-279.

Rayner, K., Carlson, A., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22,* 358-374.

Rayner, K., Garrod, S., & Perfetti, C. (1992). Discourse influences during parsing are delayed. *Cognition, 45*, 109-139.

Rolls, E. T., & Tovee, M. J. (1995). Sparseness of the neuronal representation of the stimuli in the primate visual cortex. *Journal of Neurophysiology, 73*, 713-726.

Rosch, E. (1973). Natural categories. *Cognitive Psychology, 4*, 328-350.

SAS Institute. (1989). *SAS/STAT user's guide.* Cary, NC: Author.

Schmidt, T. (2002). The finger in flight: Real-time motor control by visually masked color stimuli. *Psychological Science, 13*, 112-118.

Sheliga, B. M., Riggio, L., & Rizzolatti, G. (1995). Spatial attention and eye movements. *Experimental Brain Research, 105*, 261-275.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49*, 238-299.

Song, J., & Nakayama, K. (2006). Role of focal attention on latencies and trajectories of visually guided manual pointing. *Journal of Vision, 6*, 982-995.

Spivey, M.(2007). *The continuity of mind*. New York, NY: Oxford University Press.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1521-1543.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, 102*, 10393-10398.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of syntactic context on syntactic ambiguity resolution. *Cognitive Psychology, 45*, 447-481.

Spivey-Knowlton, M. J. (1996). *Integration of visual and linguistic information: Human data and model simulations*. Unpublished Ph. D. thesis. University of Rochester, Rochester.

Swets, B., Desmet, T., Clifton, C., Jr., & Ferreira, F. (2005). Adjunct attachment isn't just a race, either. Poster presented at the 19th Annual CUNY Conference on Human Sentence Processing, Tucson, AZ

Tabor, W. & Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*: 431-450.

Tabor, W., & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science, 23*, 491-515.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*, 355-370.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

Theeuwes, J., Olivers, C. N. L., & Chizk, C. L. (2005). Remembering a location makes the eyes curve away. *Psychological Science*, *16*, 196-199.

Tipper, S. P., Howard, L. A., & Jackson, S. R. (1997). Selective reaching to grasp: Evidence for distractor interference effects. *Visual Cognition, 4*, 1-38.

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39*, 558-592.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*, 89-134.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language, 33*, 285-318.

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences. *Journal of Memory and Language*, *55*, 335-362.

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*, 284-307.

van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against constraint-based and two-stage models. *Journal of Memory and Language, 45*, 225-258.

Ward, L. M. (2002). *Dynamical cognitive science*. Cambridge, MA: MIT Press.

Weinberg, A. (1993). Parameters in the theory of sentence processing: Minimal commitment theory goes east. *Journal of Psycholinguistic Research, 22*, 339-364.

Wolfe, J. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, *9*, 33-39.

Zadeh, L. A. (Ed.) (1975). *Fuzzy sets and their applications to cognitive and decision processes*. New York, NY: Academic Press.

Zwaan, R., & Taylor, L. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General, 135*, 1-11.

CHAPTER 4

Phonological Typicality Influences On-line Sentence Comprehension

The principle of "the arbitrariness of the sign" (de Saussure, 1916) has been a cornerstone of the study of language for over a century and is often highlighted as one of its central design features (Hockett, 1960). Except in rare cases of onomatopoeia and sound symbolism, words are considered to be arbitrary symbols that do not resemble what they stand for. Indeed, even prototypical onomatopoeia, such as animal sounds, appear highly idiosyncratic when compared cross-linguistically (Pinker, 1999). For example, the words for the noises that pigs make differ dramatically across languages, from *buubuu* in Japanese and *ut-it* in Vietnamese to *øf* in Danish, *rok-rok* in Croatian, and *oink-oink* in English. It is perhaps therefore not surprising that most modern frameworks for understanding language assume that there is little, if any, relationship between the sound of a word and how it is used (e.g., Goldberg, 2006; Jackendoff, 2002; Pinker, 1999). In this paper, however, we demonstrate that there is a systematic relationship between the sound of a word and its lexical category, and that this relationship affects language processing.

Previous research on language development has suggested that the relationship between a word's phonology and how it is used is not entirely arbitrary. For example, several phonological properties, including lexical stress (Gleitman & Wanner, 1982), number of phonemes (Morgan, Shi, & Allopenna, 1996), and vowel duration (Swanson, Leonard, & Gandour, 1992), differ between function words (determiners, prepositions, etc.) and content words (nouns, verbs, adjectives, adverbs), and newborn infants appear to be able to use such cues to differentiate these two major syntactically-motivated categories of words (Shi, Werker, & Morgan, 1999). Nouns and verbs also differ in terms of their phonological properties, and this may be important for early acquisition of syntax (Kelly, 1992; Monaghan, Chater, & Christiansen, 2005). Corpus-based analyses of child-directed speech indicate that nouns can be differentiated from verbs in terms of differences in phonological cues such as syllabic complexity, lexical stress position, and number of syllables (Durieux & Gillis, 2001; Morgan et al., 1996; Monaghan et al., 2005). Sensitivity to these cues begins early. E.g., four-day-old infants can detect differences in syllable number among isolated words (Bijeljac, Bertoncini, & Mehler, 1993), and by age three, children can use differences in number of syllables to guide their interpretation of novel words (Cassidy & Kelly, 1991). Moreover, phonological cues have also been shown to improve the learning of artificial languages by both children (Brooks, Braine, Catalano, Brody, & Sudhalter, 1993) and adults (Monaghan et al., 2005). Together, these studies indicate that nouns are distinct from verbs in terms of their phonological properties, and that children are not only sensitive to such cues but also appear to utilize them to facilitate learning.

Given the potential importance of phonological cues for syntactic development, we predicted that they would continue to play a role in adulthood as constraints on syntactic processing. Indirect support for this prediction comes from sentence production studies in which adults show sensitivity to phonological cues that may distinguish nouns from verbs. Adults are more likely to use a nonsense word as a noun when it is multi-syllabic (Cassidy, & Kelly, 2001) or has stress on the first syllable (Kelly, 1998). Here, we investigate the degree to which sensitivity to phonological cues extends to the on-line processing of sentences, focusing on the two major lexical categories of nouns and verbs.

If the phonological properties of nouns differ systematically from those of verbs, then nouns should form coherent clusters in phonological space in which nouns tend to be closer to one another than to verbs, and vice versa for verbs. We quantify the phonological clustering of nouns and verbs by measuring the distance between words within and across lexical categories. This corpus analysis shows that there exist coherent probabilistic constraints between a word's phonological form and its lexical category. Analyses of lexical naming latencies in Experiment 1 indicate that these constraints influence lexical processing, with nouns and verbs that are typical of their lexical category being accessed faster. Experiments 2 and 3 demonstrate a similar effect of *phonological typicality*—the degree to which the phonology of a given word is typical of other words in its lexical category—when processing nouns and verbs in the context of simple unambiguous sentences. Finally, Experiment 4 shows that phonological typicality directly affects on-line comprehension of sentences containing syntactic ambiguities arising from the presence of noun/verb homonyms.

Measuring Phonological Typicality

To determine the extent to which the phonological properties of words cluster together coherently within lexical categories, we extracted all the 3,158 monosyllabic nouns and verbs that were classified unambiguously according to lexical category in the CELEX database (Baayen, Pipenbrock, & Gulikers, 1995). We represented each word in terms of three slots for onset, two slots for nucleus, and three slots for the coda. Hence, the word *kelp* was represented as /k.._E._lp./ and the word *street* as /st®_ii_t../, where "." denotes an empty slot. For each word, phonemes were represented in terms of eleven phonemic features (adapted from ref. Harm & Seidenberg, 1999). We then computed the Euclidean distance between the target word and each of the nouns to measure the mean noun distance, and between the target word and each of the verbs to measure the mean verb distance. For example, for the noun /mArb´l/ the mean distance to all nouns was 8.93 whereas the distance to all verbs was 9.49, indicating that *marble* is closer in terms of its phonology to nouns

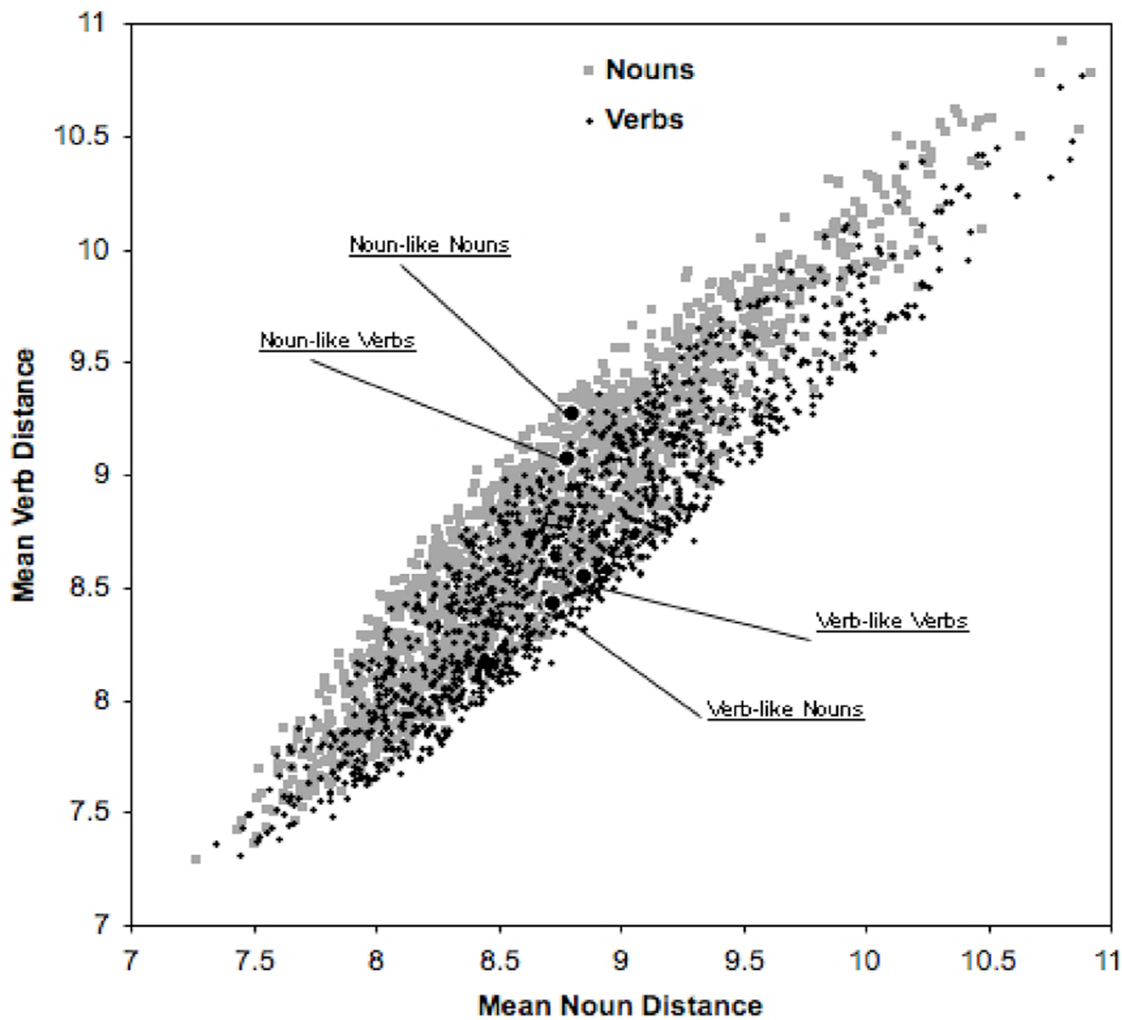

Figure 4.1. The 3,518 words from the corpus analyses in Experiment 1 plotted as a function of their mean Euclidian distance in phonological feature space to all nouns (x-axis) and all verbs (y-axis). Nouns (grey squares) tend to cluster in the upper left part and the verbs (black diamonds) in the lower right part of the figure. The points labeled Noun-like Nouns and Verb-like Nouns indicate the center of the phonologically typical and atypical nouns, respectively, used in Experiment 2. Similarly, the points Verb-like Verbs and Noun-like Verbs denote the center of the typical and atypical verbs used in Experiment 3.

than to verbs.

Each of the 1,742 nouns and 1,416 verbs in the analysis are plotted in Figure 4.1 as a function of their mean distance to all nouns and all verbs. Although there is considerable variation within each lexical category, separate clustering of nouns (upper left) and verbs (lower right) are visible in phonological space. There is, however, also a large overlap between nouns and verbs within the space, indicating that some nouns are closer overall to verbs than they are to other nouns, and, similarly, some verbs are closer to other nouns than they are to verbs. The points labeled Noun-like Nouns and Verb-like Nouns denote words that are phonologically typical and atypical of nouns, respectively. These points show the centers of the words used in Experiment 2. Similarly, the points Verb-like Verbs and Noun-like Verbs indicate, respectively, the centers of the phonologically typical and atypical words used in Experiment 3.

To test the significance of the noun and verb clusters, we performed Monte Carlo analyses in which the category labels were randomly assigned to the 3,158 words, and the same distance measures were computed. Over both nouns and verbs, words were significantly closer to other words of their own category ($p < .001$). This effect was also found when nouns and verbs were considered separately. Nouns were significantly closer to other nouns than would be expected by chance ($p < .001$), and verbs were significantly closer to other verbs than would be expected by chance ($p < .004$). These results confirm that the noun and verb clusters, discernable in Figure 1, are phonologically coherent and differ significantly from one another.

The analyses so far have involved measures of global similarity, where the phonological coherence was quantified in terms of the mean distance of a word to the remaining 3,157 nouns and verbs. We performed additional analyses to test whether coherence can also be observed locally for each individual word by testing whether the nearest neighbor to each word was of the same lexical category. For example, for *marble* the nearest neighbor in phonological space was the noun *barbel* at a distance of 2.65. When locating the word with the smallest Euclidian distance to the target word, 65.3% of the nouns had other nouns as nearest neighbors, and 64.7% of the verbs had verbs as nearest neighbors. A Monte Carlo analysis demonstrated that these results were highly significant: for nouns and verbs combined, for nouns only, and also for verbs only, p's < .001.

These coherence analyses confirmed that nouns are closer to one another than they are to verbs in terms of their phonology, and, similarly, that verbs are closer to one another than they are to nouns. These findings motivate the hypothesis that a word's phonological typicality can influence how readily it is accessed.

Experiment 1: Naming Latency Analysis

To test our hypothesis that phonological typicality should influence the processing of single words, we reanalyzed an existing database of lexical naming latencies (Spieler & Balota, 1997). We repeated the hierarchical regression analysis from the original study on the unambiguous nouns and verbs in the data set to test the extent to which phonological typicality could account for variance after other variables had been entered into the analysis. Nouns and verbs were analyzed separately.

Method

Naming Latency Data in Experiment 1

The data set (Spieler & Balota, 1997) used in the Naming Latency Analysis was produced by 31 Washington University undergraduates, who named 2,820 individually presented monosyllabic words. Several variables were found to account for portions of the variance in naming RTs (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Spieler & Balota, 1997), including features relating to the phonemic properties of the onset (e.g., dental, palatal, fricative, nasal, voiced), log-frequency, orthographic neighborhood size, and length. At the first step in our regression analyses, we entered the thirteen onset-phoneme properties (Spieler & Balota, 1997). At the second step, we entered log-frequency, orthographic neighborhood, length, imageability (from the MRC database, Coltheart, 1981), and familiarity. At the third step, for the noun analysis we entered distance from each word to all other nouns, for the verb analysis we entered distance from each word to all other verbs, and for both sets of words, we entered both distance to nouns and distance to verbs simultaneously. There were 370 nouns and 70 verbs in the analyses.

Results and Discussion

The results for nouns are shown in the upper-half of Table 4.1. The onset-phoneme coding accounted for similar variance to that found in the original analysis (Spieler & Balota, 1997) for both nouns and verbs. For nouns, log-frequency, neighborhood size, familiarity, and imageability were significant predictors of response times (RTs). For the final step, distance to nouns was a significant predictor for nouns, indicating that nouns closer to other nouns were responded to more quickly than those distant from other nouns. When both distance to nouns and distance to verbs were entered at the final step, neither was a significant predictor.

The results for verbs are shown in the lower-half of Table 4.1. For verbs, length and familiarity were significant predictors. For the final step, distance to verbs was a significant predictor, indicating that verbs phonologically similar to other verbs were responded to more quickly than verbs distant from other verbs. When both distance to nouns and distance to verbs were entered, both were significant predictors. Verbs that are closer to verbs and more distant from nouns were responded to most quickly.

Table 4.1

Regression results for Experiment 1

	Nouns			Verbs		
	ß-weight	*t* value	R^2	ß-weight	*t* value	R^2
Step 1			.296			.556
Onset-phoneme						
Step 2			.474			.654
Log-frequency	-.193	-3.237***		.344	1.853	
Length	-.216	-4.327***		.125	1.081	
Neighborhood size	.050	1.007		.300	2.277*	
Familiarity	-.154	-2.555*		-.530	-2.883**	
Imageability	-.095	-2.248*		.108	1.132	
Step 3			.482			.697
Noun distance	.101	2.346*		.265	2.683**	
Step 3			.485			.742
Noun distance	-.007	-.090		-.465	-2.967**	
Verb distance	.121	1.540		.646	4.090***	

*p < .05, **p < .01, ***p < .001

The results from Experiment 1 indicate that the typicality of a word's phonological representation influences the speed with which it can be read out aloud. This suggests that adults are sensitive to the systematic relationship between the phonology of a word and its lexical category when reading words in isolation. In Experiments 2 and 3 we test the prediction that the phonological typicality of nouns and verbs also should influence on-line processing of words in sentences.

Experiment 2: Noun Study

Experiment 2 aimed to determine whether phonological typicality would influence RTs on nouns occurring in an unambiguous syntactic structure in which a noun would be strongly expected. To produce a single measure of phonological typicality for both nouns and verbs, we subtracted the distance from a given word to all verbs from the distance from that word to all nouns. Negative values indicate that the word is closer to nouns and thus has a noun-like phonology, whereas positive values indicate that the word has a verb-like phonology because it is closer to verbs. For example, /mArb´l/ has a phonological typicality of (8.93–9.49=) -0.56, indicating that *marble* has a noun-like phonology. Based on the results of Experiment 1, we predicted that noun-like nouns would be read more quickly than verb-like nouns.

We identified 10 verbs that exhibit a strong structural bias to be followed by a noun phrase (NP). Ten sentence frames were then constructed from the NP-biased verbs (*saved*, in example 1). All words through the second determiner *the* were held constant across both sentences in each frame.

1(a) The curious young boy saved the *marble* that he found on the playground.

1(b) The curious young boy saved the *insect* that he found in his backyard.

Two sentence versions were constructed from each frame. One version included an NP with a noun-like noun (*marble,* 1a). The other version contained a verb-like noun (*insect,* 1b).[6] The sentences were presented to participants using a self-paced reading task in which the RT for each word was recorded.

Method

Participants

Twenty-two native English speakers (M=20.68 years, SD=2.03) from Cornell University participated in Experiment 2 for either $5 or extra course credit.

Materials

We selected the verb frames for the Noun Study from a prior norming study (Connine, Ferreira, Jones, Clifton, & Frazier, 1984). The mean percentage of NP completions for the verbs selected for this study was 87.7% (SD=6.8%), indicating an overwhelming structural bias to take an NP.

We controlled for several potential confounds (means and *SDs* for each control variable, per condition, appear in Table 4.2, top panel): No significant differences between the noun-like vs. verb-like target nouns existed on CELEX-based frequency, $t(18)$=.26, p=.801, orthographic length, $t(18)$=.95, p=.355, number of phonemes, $t(18)$=1.62, p=.123, or number of phonological neighbors, $t(18)$=1.42, p=.172. There were also no differences between noun-like and verb-like noun sentences in the web-based occurrence of the word triples ("trigrams") involving the frame verb, *the*, and the target noun (e.g., *saved the marble* vs. *saved the insect*), $t(18)$=.14, p=.888. We used Google-based frequencies because the occurrence of specific triples of words is quite rare even in relatively large corpora. Although web-based word co-occurrence frequencies incorporate a certain amount of noise, the resulting frequencies are not only highly correlated with corpus-based frequencies (when available), but provide

[6] All experimental sentence used in Experiments 2-4 appear in Appendix A.

even better correlations with human plausibility judgments than do corpus-based frequencies (Keller & Lapata, 2003).

Table 4.2

Means (SDs) associated with the control t-tests for Experiments 2 and 3

	Control					
Study	Frequency	Phoneme Number	Phon Neighbors	Orthographic Length	Overall Plausibility	Trigram Frequency
Experiment 2 (Nouns)						
Noun-like Nouns	546 (482.3)	6 (.67)	1.1 (.99)	6.4 (1.65)	5.74 (.69)	31892.9 (99826.88)
Verb-like Noun	642.2 (1087.3)	5.4 (.97)	3 (4.11)	5.8 (1.14)	5.7 (.58)	26074.4 (99827.55)
Experiment 3 (Verbs)						
Verb-like Verbs	494.4 (436.6)	5.7 (1.34)	2.1 (4.15)	5.9 (1.1)	5.61 (.85)	13818 (18360.54)
Noun-like Verbs	492 (435.2)	5.1 (.74)	3.9 (2.77)	5.5 (.85)	5.32 (.88)	6633.9 (14797.22)

To ensure that the sentences containing noun-like target nouns were not significantly more plausible than the sentences containing verb-like target nouns, we conducted a norming study. Twenty separate native English-speaking Cornell undergraduates rated sentences for plausibility on a seven-point Likert-type scale (7=Very Plausible). The items, along with 20 unrelated fillers, were counterbalanced across two lists. There were no significant differences in overall plausibility ratings, $t(18)=.14$, $p=.890$.

The 20 experimental sentences were counterbalanced across two different presentation lists in such a way that each list contained five noun-like noun sentences and five verb-like noun sentences, but only one version of each of the 10 frames. Each list also contained 50 unrelated filler items and eight practice items.

Procedure

Participants were randomly assigned to one of the two presentation lists. All sentences were randomly presented in a non-cumulative, word-by-word moving window format. After a brief tutorial, participants were instructed to press the 'GO' key to begin the task. The entire test item appeared on the center (left-justified) of the screen in such a way that dashes preserved the spatial layout of the sentence, but masked the actual characters of each word. As the participant pressed the 'GO' key, the word that was just read disappeared and the next one appeared. RTs (msec) were recorded for each word. After each sentence had been read, participants responded to a Yes/No comprehension question, and upon another key press, the next item appeared.

Results and Discussion

RTs on each target word were length-adjusted to eliminate differences between conditions due to character-length (Ferreira & Clifton, 1986). First, using the raw RTs on all words in both the experimental and filler items, we computed a regression equation predicting each participant's overall RT per word from the number of characters in each word. The equation was used to generate an expected RT on each word given its length. Expected RTs on each word were then subtracted from the observed RTs, and the resulting adjusted RTs used for all analyses.

Comprehension question accuracy was high: 98.2% for noun-like target noun sentences vs. 97.3% for verb-like target noun sentences. However, as illustrated in Figure 4.2 (left panel), the noun-like nouns were processed significantly faster than the verb-like nouns, $t(21)=2.84$, $p=.01$.[7] Given that it has been suggested that differences in the number of syllables may affect whether a word is more likely to be perceived as

[7] The combination of the tightly controlled stimuli in Experiments 2-4, and their counterbalancing across conditions, makes item analyses inappropriate (Raaijmakers, Schrijnemakers, & Gremmen, 1999)

a noun or a verb—with multiple syllables being indicative of a noun (Cassidy & Kelly, 1991)—we conducted a second RT analysis in which we factored out syllable number using the same regression-based length-adjustment procedure as before, and observed a commensurate significant difference, $t(21)=3.71$, $p=.001$. The faster responses for noun-like compared to verb-like nouns indicate that adults are sensitive to the typical phonological properties of words in the lexical category of nouns. In the next experiment, we investigate whether a similar sensitivity can be found for the lexical category of verbs.

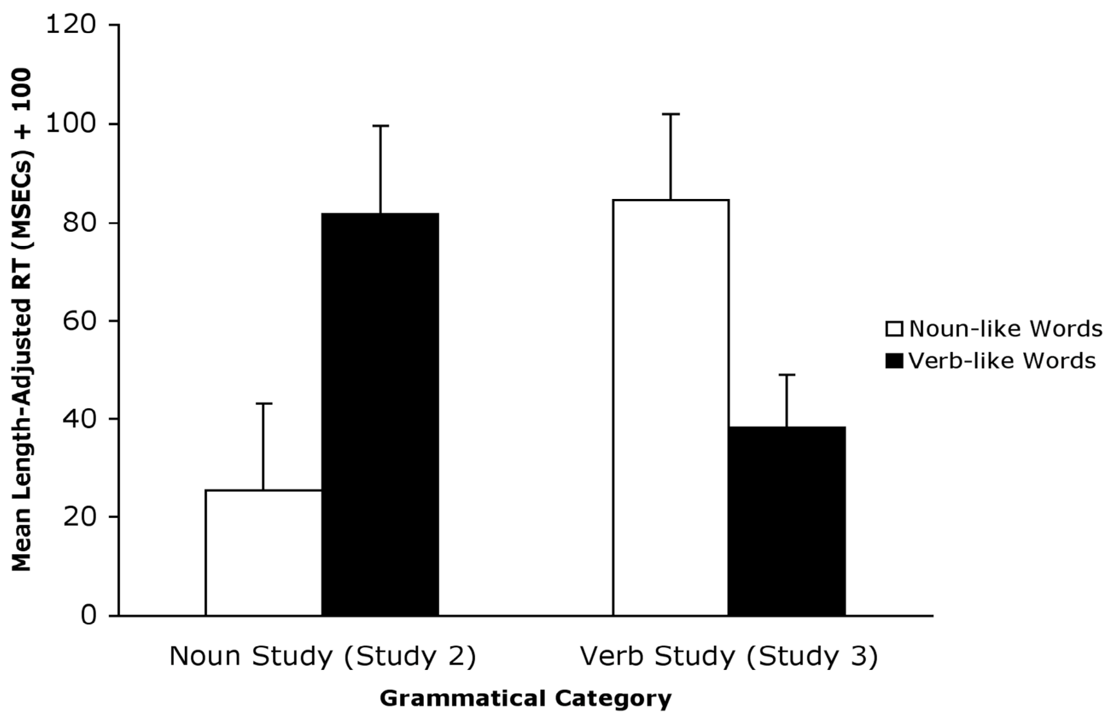

Figure 4.2. Mean RTs (and standard errors) for the phonologically typical and atypical conditions in Experiments 2 and 3. After length-adjustment, a constant of 100 was added to make the figure easier to interpret.

Experiment 3: Verb Study

Experiment 3 was designed to determine whether the effect of phonological typicality on RTs in unambiguous sentences would extend to verbs. We predicted that verb-like verbs would be read faster than noun-like verbs.

We identified 10 verbs exhibiting a strong tendency to take an infinitival complement (*inf-comp*) structure (e. g., ... *is trying to*...). Ten sentence frames were then constructed from the chosen frame verbs (*tried,* in example 2). Two versions of each frame were constructed in which all words up through the infinitival *to* marker were held constant across both sentences in each frame.

2(a) The young girl had tried to *amuse* herself while waiting for her mother by working on a crossword puzzle.

2(b) The young girl had tried to *ignore* the boy that kept on pulling her hair during recess.

One version included an *inf-comp* structure with a verb-like target verb (*amuse*, 2a). The other version included an *inf-comp* with a noun-like target verb (*ignore,* 2b).

Method

Participants

Twenty-two native English speakers (*M*=20.46 years, *SD*=1.47) from Cornell University participated in Experiment 2 for either $5 or extra course credit.

Materials

For the Verb Study, four verbs were selected from prior norming data (Connine et al., 1984), for which over 80% of participants followed the verb with an *inf-comp* when asked to use the verb in a sentence. The other six verbs were selected from a norming study in which we presented 15 separate native English speakers from Cornell University with a sentence completion task containing 13 sentence stems that ended with the verb of interest (e.g., *The employees were expected…*). Participants were asked to complete each stem with whatever sounded most natural to them. Two of the verbs selected for the study elicited 100% *inf-comp* completion, and the other four elicited 95% *inf-comp* completion.

The means and *SD*s for each control variable, per condition, appear in Table 4.2, bottom panel. There were no significant differences between the verb-like and noun-like verbs on CELEX-based overall frequency, $t(18)=.01$, $p=.990$, number of nearest phonological neighbors, $t(18)=1.14$, $p=.269$, orthographic length, $t(18)=.91$, $p=.375$, the number of phonemes, $t(18)=1.24$, $p=.230$, or their occurrence in trigrams consisting of the frame verb, *to*, and the target verb (e.g., *tried to amuse* vs. *tried to ignore*), $t(18)=.96$, $p=.348$. Additionally, 20 separate Cornell undergraduates participated in a plausibility norming study using the same method as in Experiment 2. There were no significant differences in plausibility between the sentences containing verb-like and noun-like verbs, $t(18)=.75$, $p=.462$. The materials were counterbalanced and presented as described in Experiment 2.

Results and Discussion

Again, comprehension accuracy was high: 98.2% correct for verb-like verb sentences vs. 95.5% correct for noun-like verb sentences. As illustrated in Figure 4.2 (right panel), however, the verb-like verbs were processed significantly faster than the noun-like verbs, $t(21)=3.15$, $p=.005$. The syllable length-adjusted RT analyses also yielded a significant difference, $t(21)=2.86$, $p=.009$. These results indicate that participants were sensitive to the phonological typicality of verbs. Participants took longer to read the verbs that were more typical of nouns in terms of their phonology.

One possible concern with Experiments 2 and 3 is that orthographic regularities—instead of phonological typicality—could be the cause of the observed difference in RTs. To address this concern, we created a measure of orthographic typicality that directly parallels phonological typicality using Coltheart's N (Coltheart, Davelaar, Jonasson, & Bresner, 1977), by subtracting the number of verbs that can be found by changing one letter in the target word from the number of nouns generated by the same process of single-letter modification. This measure of orthographic typicality was then used to predict RTs on each target word in a regression equation. We found that orthographic typicality did not predict length-adjusted RTs on the target words for Experiments 2 and 3, $t(19)=1.02$, $p=.323$, $t(19)=.70$, $p=.496$, respectively. To further address concerns about orthographic typicality, we controlled for it *a priori* in Experiment 4.

The results of Experiments 2 and 3 showed that when the phonological typicality of a word is not congruent with the expected lexical category of the word, on-line processing is, at least momentarily, impeded. This effect is robust for both nouns and verbs, demonstrating on-line effects of phonological typicality on unambiguous sentences. To determine whether the systematic phonological regularities of nouns and verbs also affect sentence interpretation, Experiment 4 investigates whether phonological typicality can influence on-line parsing preferences during the processing of syntactically ambiguous sentences.

Experiment 4: Homonym Study

We investigated the influence of phonological typicality on the processing of syntactic ambiguities arising from the lexical category ambiguity associated with noun/verb (N/V) homonyms. A classic example of this type of ambiguity can be seen in the sentence fragment *I know that the desert trains…* (Frazier & Rayner, 1987; MacDonald, 1993), in which the lexical ambiguity of the homonym *trains* introduces a syntactic ambiguity with respect to the continuation of the sentence. A noun reading would lead to the expectation of an upcoming verb (as in, *...could resupply the camps*) and a verb reading would result in the expectation of some type of complement (as in, *...soldiers to be tough*). We hypothesized that the phonological typicality of the N/V homonym would have an on-line influence on whether participants would expect a verb or complement continuation of the sentence. Specifically, we predicted that noun-like N/V homonyms would cause participants to experience processing difficulties when the sentence was resolved with a verb interpretation of the N/V homonym, and vice versa for verb-like N/V homonyms.

Twenty sentence frames incorporating a syntactic ambiguity arising from a N/V homonym were constructed consistent with the previous example.

3 (a) Chris and Ben are glad that the bird *perches* seem easy to install.
(b) Chris and Ben are glad that the bird *perches* comfortably in the cage.
4 (a) The teacher told the principal that the student *needs* were not being met.
(b) The teacher told the principal that the student *needs* to be more focused.

Ten sentence frames contained a noun-like N/V homonym, such as *perches* in (3), and 10 contained a verb-like N/V homonym, such as *needs* in (4). Two different versions of each sentence frame were constructed; one version contained a noun resolution of the syntactic ambiguity, as in sentences (3a) and (4a), whereas the other contained a verb resolution of the ambiguity, as in (3b) and (4b). Across all 40 sentences, the N/V homonym occupied the ninth word position, followed by four words.

Method

Participants

Forty native English speakers (*M*=20.1 years, *SD*=1.15) from Cornell

University participated in Experiment 2 for either $5 or extra course credit.

Materials

Because the Homonym Study involved a syntactic manipulation, we controlled for stimulus-specific factors that may influence syntactic processing, namely frequency and plausibility. The means and *SD*s for each control variable, per condition, appear in Table 4.3. There was no significant difference for the noun-like

Table 4.3

Means (SDs) associated with the control t-tests for Experiment 4

	Control				
	Freq. of Use as a Noun	Freg. of Use as a Verb	Orthographic Typicality	Noun Compound Plausibility Rating	Overall Plausibility Rating
Noun-like N/V Homonyms	51% (18.1)	49% (18.1)	2.15 (1.67)	4.22 (.58)	
- Noun-Resolved					5.24 (.81)
- Verb-Resolved					4.76 (.70)
Verb-like N/V Homonyms	53.8% (8)	46.2% (8)	2.8 (1.75)	4.33 (.33)	
- Noun-Resolved					5.16 (.71)
- Verb-Resolved					5.00 (.65)

N/V homonyms in the frequency of usage as a noun vs. as a verb, $t(9)=.17$, $p=.87$, nor was there a difference for the verb-like N/V homonyms, $t(9)=1.54$, $p=.15$. Additionally, we used web-based frequency counts to ensure that the trigrams involving the potential noun compound and the disambiguating word (e.g., *bird perches seem* vs. *bird perches comfortably*) were not more frequent for noun resolutions than for verb resolutions in both noun-like, $t(18)=.90$, $p=.381$, and verb-like, $t(18)=1.01$, $p=.328$, homonym sentences. Likewise, the trigrams involving the ambiguous homonym and the two following disambiguating words (e.g., *perches seem easy* vs. *perches comfortably in*) were not more frequent for either resolution in the noun-like, $t(18)=1.00$, $p=.333$, or verb-like, $t(18)=1.00$, $p=.331$, homonym items. Finally, we controlled for orthographic typicality to ensure that it did not differ from chance for the noun-like, $t(9)=.80$, $p=.496$, or the verb-like N/V homonyms, $t(9)=1.41$, $p=.191$.

Two different plausibility norming studies were conducted on the materials. First, to ensure that noun compounds sounded equally plausible when involving either noun-like or verb-like homonyms (Haskell, MacDonald, & Seidenberg, 2003), we presented 20 separate Cornell students with the 20 noun compounds used in this study, along with 30 filler items. They were asked to indicate, on a seven-point Likert-type scale, how likely the compound was to be a noun compound. The noun-like N/V homonym compounds were not rated differently than the verb-like N/V homonym compounds, $t(19)=1.07$, $p=.297$.

Second, we presented 20 separate Cornell undergraduates with one of two counterbalanced lists containing half of the noun-like and half of the verb-like N/V homonym items, in their complete form, intermixed with 16 filler items, and asked them to rate the overall plausibility of each sentence on a seven-point Likert-type scale. We found no significant difference in overall plausibility ratings between the noun and verb resolutions for the noun-like N/V homonym items, $t(18)=1.41$, $p=.175$, and none between the noun and verb resolutions for the verb-like N/V homonym items, $t(18)=.53$, $p=.605$.

The 40 sentences were counterbalanced across two different presentation lists such that each participant saw five sentences in each possible condition, but only one version of each of the 20 sentence frames. The items were presented along with 30 unrelated filler items and eight practice items.

Results and Discussion

Participants encountered a syntactic ambiguity upon reading the N/V homonym, which could be parsed as a noun that is modified by the preceding word to form a noun compound, or which could be interpreted as a verb. All sentences were disambiguated by the word following the N/V homonym (the 10th word). However, given some concern about the actual disambiguation point[8], we also included the 11th word. Accordingly, two segments were created, the point of ambiguity (word 9), and the point of disambiguation (consisting of words 10 and 11 averaged together).

A 2 (noun-like vs. verb-like N/V homonym) x 2 (noun vs. verb resolution) x 2 (ambiguity vs. disambiguation) repeated-measures ANOVA yielded a statistically reliable three-way interaction by-subjects, $F_1(1, 39)=19.79$, $p<.0005$, $MSE=7667.22$, and by-items, $F_2(1,18)=13.30$, $p=.002$, $MSE=2852.62$. This three-way interaction was also significant in the syllable length-adjusted analysis, $F_1(1, 39)=17.31$, $p<.0005$, $MSE=6988.21$; $F_2(1,18)=9.41$, $p=.007$, $MSE=3215.02$. Figure 4.3 illustrates the mean of the difference scores between the point of disambiguation and point of ambiguity (disambiguation minus ambiguity) for each of the four possible conditions. In all conditions, RTs increased from the point of ambiguity to the point of disambiguation. However, for sentences containing noun-like N/V homonyms, RTs increased significantly more for the verb-resolved than for the noun-resolved sentences, $t(39)=2.50$, $p=.017$. Similarly, for sentences containing verb-like N/V homonyms, RTs increased significantly more from ambiguity to disambiguation for the noun-resolved sentences than they did for the verb-resolved sentences, $t(39)=4.17$, $p<.0005$.

---

[8] In a few cases, there is a third interpretation of the ambiguity. There is a small chance that the second noun in the noun compound (e.g., *needs*, as in *the student needs*) could be considered a modifier for an upcoming head noun. However, plural nouns are rarely modifiers in English (MacDonald, 1993; see also Haskell et al., 2003).

The RT interaction demonstrates that phonological typicality can bias readers to entertain one interpretation of the ambiguity over the other. The effect of phonological typicality on processing is further illustrated, off-line, by the pattern of comprehension accuracy rates. For the noun-like N/V homonym sentences, accuracy rates were 99.5% correct on the noun-resolved sentences and 95% on the verb-resolved sentences. For the verb-like N/V homonym sentences, accuracy rates were 94.5% correct on the verb-resolved sentences and 91.5% on the noun-resolved sentences. Notably, participants were significantly more accurate on conditions where a match existed between the phonological typicality of the N/V homonym and the resolution of the sentence (*M*=9.7 correct, *SD*=.52) than on sentences containing a mismatch (*M*=9.33, *SD*=.92), $t(39)=2.49$, $p=.017$.

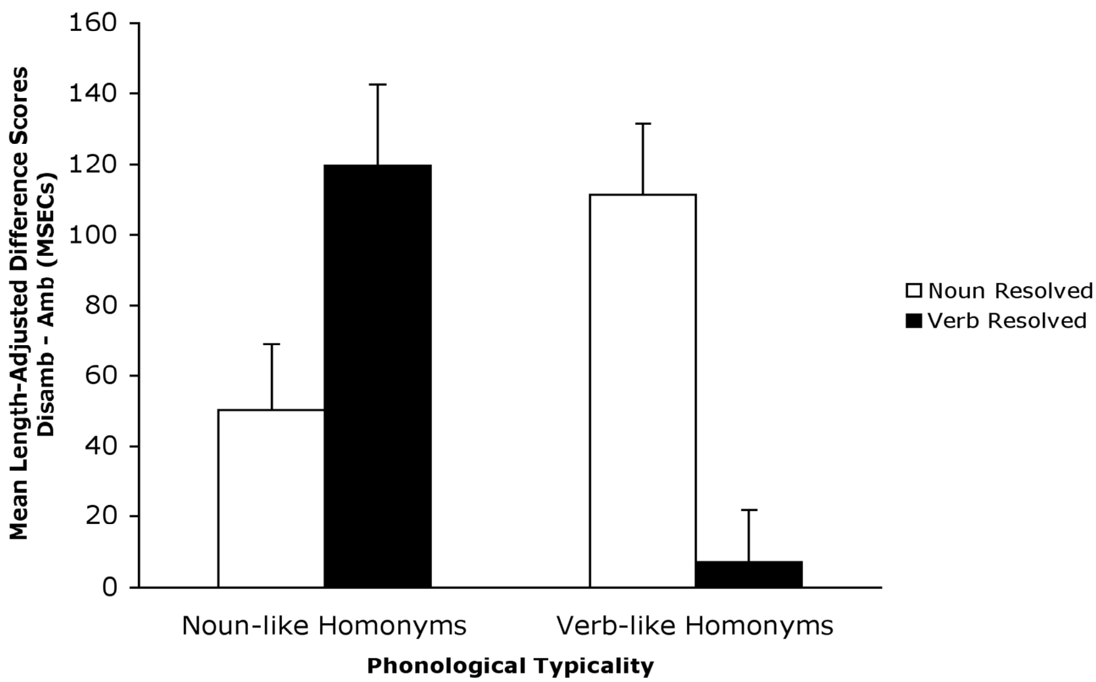

Figure 4.3. Mean difference (disambiguation minus point of ambiguity) scores (and standard errors) for each of the four possible conditions in Experiment 4. Rising bars indicate that RTs increased from the point of ambiguity to disambiguation.

In summary, not only does phonological typicality appear to bias the on-line interpretation of a syntactically ambiguous sentence, as demonstrated by the RT data, but it also influences, off-line, whether or not people eventually comprehend the sentence correctly.

General Discussion

Although it has long been known that both phonological information (Huey, 1908) and grapheme-phoneme correspondence (Gibson, Pick, Osser, & Hammond, 1962) can affect reading performance, the studies presented here are the first to demonstrate that the relationship between phonology and lexical categories can directly affect on-line language processing. Previous studies have indicated that adults are sensitive to gross-level phonological properties, such as stress (Kelly, 1998) and syllable length (Cassidy & Kelly, 1991), when producing sentences using nonsense words. In contrast, our results reveal that the more subtle phonological properties that comprise phonological typicality relative to lexical categories have an effect on both lexical and sentential processing. The corpus analysis revealed a systematic relationship between the sound of a word and whether it is used as a noun or a verb. The subsequent four experiments demonstrated that adults are sensitive to such phonological typicality both when reading isolated words aloud and when comprehending ambiguous and unambiguous sentences. Thus, contrary to what would be expected given the Saussurean principle of the “arbitrariness of the sign” (de Sausssre, 1916) our results show that the sound of a word does provide an indication of what it refers to; specifically, whether it refers to a noun or a verb.

Analyses of languages such as Dutch, French, Japanese, Mandarin, and Turkish (Durieux & Gillis, 2001; Morgan et al., 1996; Onnis & Christiansen, 2005—see Kelly, 1988, for a review) suggest that phonological cues to lexical categories are not unique to English but may be a universal property of language. Additionally, more fine-grained phonologically-based subdivisions of words within lexical categories may also be found in the form of sound symbolism (see Nuckolls, 1999, for a review). For example, *gl-* in English tends to occur in words relating to sound and vision: *glimmer, glisten, glitter, gleam, glow, glint*, etc; and people are sensitive to these sound-meaning relations as evidenced by priming experiments (Bergen, 2004). Although it is often assumed that the presence of sound symbolism would require that words with similar referents have the same phonological form across different languages (de Saussure, 1916; Pinker), we suggest that systematic relationships between sound and word use are more likely to be specific to individual languages. Indeed, phonological cues to lexical categories vary considerably across languages (Onnis & Christiansen, 2005), and we would expect more fine-grained cues to show similar cross-language variation—though some overlap may be expected due to historical relationships between languages. Each language is hypothesized to have its own constellation of phonological cues relevant for distinguishing between lexical categories and perhaps some subdivisions within these. What is important is that the cues form a reasonably coherent system within a language. However, computational simulations involving artificial neural network models learning mappings between pseudo-phonological word forms and pseudo-meanings have suggested that a considerable degree of arbitrariness in the form-meaning mappings is likely to remain important for language learning (Gasser, 2004). Crucially, these simulations indicate that from a computational perspective, a language is most easily learned if it coheres with phonological typicality in relation to lexical categories but maintains, as much as possible, arbitrary form-meaning relations.

An important implication of our results is that non-syntactic information—even in the form of phonological cues—can exert an early influence on sentence comprehension. Further investigations will be needed to establish the exact time-course within which phonological typicality may be influencing the comprehension process. However, an early effect of phonological typicality appears likely given the growing number of event-related brain potential studies indicating that the language system generates fast, probabilistic expectations for various characteristics of upcoming words, including their specific lexical category (Hinojosa, Moreno, Casado, Munoz, & Pozo, 2005) and onset phoneme (DeLong, Urbach, & Kutas, 2005). Moreover, not only does phonology facilitate the integration of word meaning with sentential context in silent reading independent of orthography (Newman & Connolly, 2004), but also, in the form of prosody, has an immediate influence on syntactic interpretation (Steinhauer, Alter, & Friederici, 1999)—even when words are presented visually (Steinhauer & Friederici, 2001) similar to Experiments 2-4.

More broadly, our results are consistent with a view of language comprehension in which the use of multiple constraints in adult sentence processing emerges as the product of a developmental process driven by the integration of multiple cues (Bates & MacWhinney, 1987; Seidenberg & MacDonald, 1999; Snedeker & Trueswell, 2004). Because language comprehension is a complex task that involves constructing an incremental interpretation of a rapid sequence of incoming words before they fade from immediate memory, the adult comprehension system has been developed to exploit multiple sources of information to facilitate the task (MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995). Many factors, including referential context (Altmann & Steedman, 1988), lexically-based verb biases (Trueswell, Tanenhaus, & Kello, 1994), and prosody (Snedeker & Trueswell, 2003), appear to constrain how an incoming string of words is processed.

Sensitivity to each of these constraints emerges gradually, following different time-scales, during language development due to relative differences in saliency and reliability. Owing to the higher reliability of lexico-syntactic contingencies, sensitivity to local word-specific cues such as phonological typicality are likely to appear earlier in children's language comprehension than the ability to use more complex cues deriving from global information sources such as referential context and prosody. We suggest that the effects of phonological typicality observed here in adult sentence processing are due to the role of phonology in the early development of lexical representations. Thus, the importance of phonological cues in language acquisition can be observed in adulthood as the influence of phonological typicality on sentence comprehension.

APPENDIX A

Experimental Sentences — Experiment 2 (Noun Study)

Note: The **bolded** noun in each sentence is the target noun of interest. The first sentence in each pair contains the noun-like noun, and the second sentence in each pair contains the verb-like noun.

The curious young boy saved the **marble** that he found on the playground.
The curious young boy saved the **insect** that he found in his backyard.

The very little girl imitated the **laughter** of the old woman with a tone of mockery.
The very little girl imitated the **infant** as soon as it began to cry.

The very angry man described the **neighbor** as a menace to society.
The very angry man described the **theft** to the policeman soon after it had occurred.

The group of friends discussed the **movie** that they had just gone to see.
The group of friends discussed the **scenes** from the movie that they found most humorous.

The terrible car accident blocked many **drivers** from the main entrance to the shopping mall.
The terrible car accident blocked many **lanes** of the town's only major highway.

The extremely generous woman bought her **daughter** many expensive gifts for her birthday.
The extremely generous woman bought her **friends** dinner in celebration of her promotion at work.

The quiet college student read the **bible** during times of intense stress.
The quiet college student read the **text** assigned by his history professor.

The conservative political commentator criticized the **lawyers** for agreeing to defend the cold-blooded killer.
The conservative political commentator criticized the **airlines** for not thoroughly screening all passenger bags.

The company truck driver unloaded the **cargo** from his truck onto the loading dock.
The company truck driver unloaded the **trunks** from his truck into his client's office.

The moving company employees carried the **sofa** from the van into the house.
The moving company employees carried the **chest** from the van into the house.

Experimental Sentences — Experiment 3 (Verb Study)

Note: The **bolded** verb in each sentence is the target verb of interest. The first sentence in each pair contains the verb-like verb, and the second sentence in each pair contains the noun-like verb.

The very old man attempted to **assist** his elderly wife with the household cleaning.
The very old man attempted to **vary** his daily routine by starting to exercise in the morning.

The town residents had continued to **await** news about the possibility of a tornado touching down close by.
The town residents had continued to **bury** people at the cemetery even though it was extremely crowded.

The young girl had tried to **amuse** herself while waiting for her mother by working on a crossword puzzle.
The young girl had tried to **ignore** the boy that kept on pulling her hair during recess.

The late student was required to **explain** the reason for her tardiness to the teacher.
The late student was required to **suffer** through detention as punishment for her tardiness.

The city government was prompted to **adopt** new laws to combat the littering problem.
The city government was prompted to **cancel** its weekly meeting due to the mayor's illness.

The juvenile offender was ordered to **attend** counseling sessions four times a week.
The juvenile offender was ordered to **follow** his probation guidelines very carefully.

The office secretary was instructed to **arrange** all of the files in alphabetical order.
The office secretary was instructed to **manage** the office while her boss was out of town.

The frugal woman was hesitant to **lend** her friend the large amount of money.
The frugal woman was hesitant to **borrow** money because she didn't want to pay the interest.

The presidential candidate was expected to **respond** to the allegations at the press conference.
The presidential candidate was expected to **alter** the nature of his campaign.

The overweight child was encouraged to **include** more vegetables in his diet.

The overweight child was encouraged to **wander** around the park in order to get more exercise.

Experimental Sentences — Experiment 4 (Noun/Verb Homonym Study)

Note: The noun/verb homonym (the point of ambiguity) in each sentence pair appears in **bold**. The two words following the noun/verb homonym (the point of disambiguation) appear in *italics*. The first sentence in each sentence-pair contains the noun resolution, and the second sentence in each pair contains the verb resolution.

“Noun-like” Noun/Verb Homonym Sentences

The jewelry dealer had explained that the diamond **sparkles** *were evidence* of authenticity.
The jewelry dealer had explained that the diamond **sparkles** *for many* different reasons.

It is quickly becoming apparent that the fire **blazes** *are getting* too intense.
It is quickly becoming apparent that the fire **blazes** *even during* severe thunderstorms.

The detectives were all pleased that the female **faces** *were easy* to identify.
The detectives were all pleased that the female **faces** *a long* jail sentence.

Chris and Ben are glad that the bird **perches** *seem easy* to install.
Chris and Ben are glad that the bird **perches** *comfortably in* the cage.

The new mother often notices that the baby **bounces** *are indicative* of happiness.
The new mother often notices that the baby **bounces** *happily in* her crib.

The meticulous usher had noticed that the guest **passes** *were no* longer valid.
The meticulous usher had noticed that the guest **passes** *food to* her boyfriend.

The experienced carpenters had observed that the cabinet **latches** *were not* properly installed.
The experienced carpenters had observed that the cabinet **latches** *improperly when* slammed shut.

Even the inexperienced bartender knew that the drunk **stammers** *were signs* of intoxication.
Even the inexperienced bartender knew that the drunk **stammers** *into the* bar daily.

The consumer had been informed that the carwash **waxes** *were cheap* and ineffective.
The consumer had been informed that the carwash **waxes** *your car* for free.

I think it is unfortunate that the government **counsels** *make very* bad decisions.
I think it is unfortunate that the government **counsels** *key witnesses* to lie.

"Verb-like" Noun/Verb Homonym Sentences

The pilots expressed some concern that the helicopter **sounds** *annoy the* suburban residents.
The pilots expressed some concern that the helicopter **sounds** *too unstable* to fly.

The radio broadcaster was told that the station **alerts** *won't be* aired today.
The radio broadcaster was told that the station **alerts** *the townspeople* to danger.

The truck driver wasn't aware that the road **bends** *were sharp* and dangerous.
The truck driver wasn't aware that the road **bends** *around the* mountain top.

The primatologist had previously written that the monkey **calls** *indicated a* potential threat.
The primatologist had previously written that the monkey **calls** *for food* when hungry.

The fisherman was well aware that the whale **dives** *could capsize* his boat.
The fisherman was well aware that the whale **dives** *near his* fishing boat.

Polly and Cindy had observed that the bear **hunts** *violated many* animal rights.
Polly and Cindy had observed that the bear **hunts** *for food* near campsites.

The young babysitter had known that the infant **smiles** *were a* good sign.
The young babysitter had known that the infant **smiles** *after it* is fed.

The teacher told the principal that the student **needs** *were not* being met.
The teacher told the principal that the student **needs** *to be* more focused.

The circus worker had mentioned that the whip **snaps** *frighten the* vicious lion.
The circus worker had mentioned that the whip **snaps** *loudly when* it's used.

The old farmer was aware that the pig **smells** *offend the* nearby neighbors.
The old farmer was aware that the pig **smells** *really bad* after eating.

REFERENCES

Altmann, G. T. M., & Steedman, M. J. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191-238.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *CELEX lexical database*. Philadelphia, PA: Linguistic Data Consortium.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition for single syllable words. *Journal of Experimental Psychology: General, 133*, 283-316.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3-73). Cambridge, UK: Cambridge University Press.

Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language, 80*, 290-311.

Brooks, P. J., Braine, M. D., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32*, 76-95.

Bijeljac, R., Bertoncini, J., & Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances. *Developmental Psychology, 2*, 711-721.

Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language, 30*, 348-369.

Cassidy, K. W., & Kelly, M. H. (2001). Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin and Review, 8*, 519-523.

Coltheart, M. (1981). The MRC psycholingusitic database. *Quarterly Journal of Experimental Psychol*ogy, *33A*, 497-505

Coltheart, M., Davelaar, E., Jonasson, J. T., & Bresner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance, VI* (pp. 535-555). Hillside, NJ: Erlbaum.

Connine, C., Ferreira, F., Jones, C., Clifton, C., & Frazier, L. (1984). Verb frame preferences: Descriptive norms. *Journal of Psycholinguistic Research, 13*, 307-319.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117-1121.

Durieux, G., & Gillis, S. (2001). Predicting grammatical classes from phonological cues: An empirical test. In J. Weissenborn & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition* (Vol. 1, pp.189-229). Amsterdam: John Benjamins.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.

Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language, 26*, 505-526.

Gasser, M. (2004). The origin of arbitrariness in language. . In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 434-439). Mahwah, NJ: Lawrence Erlbaum.

Gibson, E. J., Pick, A., Osser, H., & Hammond, M. (1962). The role of grapheme-phoneme correspondence in the perception of words. *American Journal of Psychology, 75,* 554-570.

Gleitman, L., & Wanner, E. (1982). The state of the state of the art. In. E. W. Wanner

& L. Gleitman (Eds.), *Language Acquisition: The State of the Art* (pp. 3-48). Cambridge: Cambridge University Press.

Goldberg, A. (2006). *Constructions at work: The nature of generalizations in language*. New York: Oxford University Press.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*, 491-528.

Haskell, T. R., MacDonald, M. C., & Seidenberg, M. S. (2003). Language learning and innateness: Some implications of compounds research. *Cognitive Psychology, 47,* 119-163.

Hinojosa, J. A., Moreno, E. M., Casado, P., Muñoz, F., & Pozo, M. A. (2005). Syntactic expectancy: and event-related potentials study. *Neuroscience Letters, 378*, 34-39.

Hockett, C. F. (1960). The origin of speech. *Scientific American, 203*, 89-96.

Huey, E. B. (1908). *The psychology and pedagogy of reading.* New York, NY: Macmillian.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* New York: Oxford University Press.

Keller, F., & Lapata, M. (2003). Using the web to obtain frequencies for unseen bigrams. *Computational Linguistics, 29*, 459-484.

Kelly, M. (1988). Rhythmic alternations and lexical stress differences in English. *Cognition, 30*, 107-137.

Kelly, M. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99*, 349-364.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language, 32*, 692-715.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676-703.

Monaghan, P., Chater, N., & Christiansen, M. H. (in press). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition, 96*, 143-182.

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of grammatical categories. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 263-283). Mahwah, NJ: Lawrence Erlbaum Associates.

Newman, R. L., & Connolly, J. F. (2004). Determining the role of phonology in silent reading using event-related potentials. *Cognitive Brain Research, 21*, 94-105

Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology, 28*, 225-252.

Onnis, L., & Christiansen, M. H. (2005). Happy endings for absolute beginners: Psychological plausibility in computational models of language acquisition. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 1678-1683). Mahwah, NJ: Lawrence Erlbaum.

Pinker, S. (1999). *Words and rules: The ingredients of language.* New York: Harper Collins.

de Saussure, F. (1916) *Cours de Linguistique Generale*. Paris, France: Paycot.

Seidenberg, M., & MacDonald, M. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23*, 569-588.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 72*, B11-B21.

Snedeker, J., & Trueswell, J. C. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language,* *48*, 103-130.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49*, 238-299

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8*, 411-416.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*, 191-196.

Swanson, L., Leonard, L., & Gandour, J. (1992). Vowel duration in mother's speech to young children. *Journal of Speech and Hearing Research, 35,* 617-625.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In P. D. Eimas & J. L. Miller (Eds.), *Handbook in perception and cognition, volume 11: Speech, language, and communication* (pp.217-262). San Diego, CA: Academic Press.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating the effects of lexical preference from garden-paths. *Journal of Memory and Language: Learning, Memory, and Cognition, 19*, 528-553. -319.

CHAPTER 5

Phonological Typicality Influences Sentence Processing in Predictive Contexts: A Reply to Staub et al. (2009)

Preface

Recently, Staub, Grant, Clifton, and Rayner (2009) have questioned the statistical reliability of the results from Experiments 2 and 3 in Farmer, Christiansen, and Monaghan (2006). In an eye-tracking experiment and a separate self-paced reading experiment, they attempted to replicate the phonological typicality effect in the unambiguous noun study (when expecting an NP, noun-like nouns are read more quickly than verb-like nouns) and the unambiguous verb study (verb-like verbs are read more quickly than noun-like verbs when expecting an infinitival complement). However, there were a series of systematic differences between the studies of Staub et al. and the original Farmer et al. experiments. Below, Farmer, Monaghan, Misyak, and Christiansen (*submitted*) detail one of the most glaring differences between the two sets of studies, and provide new experimental evidence supporting the notion that Staub et al.'s alteration of the original studies more than likely contributed to their inability to replicate the original results.

Introduction

Language comprehension is a complex task that involves constructing an incremental interpretation of a rapid sequence of incoming words before they fade from immediate memory, and yet the task is typically carried out efficiently and with little conscious effort. In order to achieve this level of speed and efficiency, the adult comprehension system exploits multiple sources of information that might facilitate the task. Many factors, including referential context (e.g., Altmann, van Nice, Garnham, & Henstra, 1998), lexically-based verb biases (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997), and prosody (e.g., Snedeker & Yuan, 2008) appear to constrain how an incoming string of words is processed (see Altmann, 1998; Elman, Hare, & McRae, 2004, for reviews). Such informative cues are not only used to resolve previously encountered ambiguous input, but also to generate syntactic expectations for what may come next. Indeed, a growing number of studies suggest that prediction-based processing is a necessary component of efficient and effortless interpretation of language as it unfolds in time (e.g., Altmann et al., 1998; Rayner, Ashby, Pollatsek, & Reichle, 2004; Staub & Clifton, 2006; see Hagoort, in press; Pickering & Garrod, 2007, for reviews).

Convergent results have been found in event-related potential (ERP) experiments, showing that highly specific expectations are generated for both lexical-category and phonological properties of upcoming words given a predictive context. Thus, during on-line sentence processing, context-based expectations are rapidly generated for (a) the grammatical gender of upcoming words, such as specific gender markings of nouns following a gender-marked adjective in spoken Dutch (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005) or a gender-marked adjective in written Spanish (Wicha, Moreno, & Kutas, 2004), (b) the lexical category of the next word (e.g., a noun following a determiner, Hinojosa, Moreno, Casado, Muñoz, & Pozo, 2005), and (c) the onset phoneme of the next word (e.g., words starting with a consonant after '*a*' or a vowel after '*an*' in English, DeLong, Urbach, & Kutas, 2005).

Building on this work, Farmer, Christiansen, and Monaghan (2006) investigated whether phonological typicality—the degree to which the sound properties of an individual word are typical of other words in its lexical category—influences on-line language processing, testing a hypothesis originally put forward by Kelly (1992) and supported by recent work on language acquisition (e.g., Cassidy & Kelly, 2001; Fitneva, Christiansen, & Monaghan, in press; Monaghan, Christiansen, & Chater, 2007). Farmer et al. presented results from a corpus analysis, showing that nouns tend to sound like other nouns and verbs like other verbs; that is, nouns and verbs form separate coherent, yet partially overlapping, clusters in phonological space. Thus, some words are more typical in their phonology of their respective lexical class than others. Farmer et al. referred to words that are typical, in terms of their phonology, of the class of nouns as 'noun-like,' and words more phonologically typical of verbs as 'verb-like'. They then reported four experiments demonstrating the impact of such phonological typicality on the processing of nouns and verbs. Using a self-paced reading methodology, two of the experiments focused on the processing of unambiguous sentences and elicited significant effects of phonological typicality. One experiment involved sentence frames designed to strongly predict that a noun will come next, whereas the frames in the other experiment were created to generate strong expectations for a verb. When the preceding context generated a strong expectation for an upcoming noun, noun-like nouns were read faster than verb-like nouns, and when the context was highly predictive of a verb, verb-like verbs were read faster than noun-like verbs.

In a subsequent study, Staub, Grant, Clifton and Rayner (in press) failed to find effects of phonological typicality in experiments examining eye-tracking and self-paced reading times when they combined the unambiguous noun and verb materials from Farmer et al.'s two separate experiments. Staub et al. interpreted their null results as indicating that phonological typicality may not influence normal reading. However, in the current study, we demonstrate that the replication failure is due to an unforeseen consequence of Staub et al.'s interleaved design, and that when this design characteristic is accounted for, the effect of phonological typicality re-emerges.

Consider the following examples of the experimental sentences:

(1a) The curious young boy saved the *marble* that he … (Noun-like Noun)

(1b) The curious young boy saved the *insect* that he … (Verb-like Noun)

(2a) The very old man attempted to *assist* his elderly wife ... (Verb-like Verb)

(2b) The very old man attempted to *vary* his daily routine ... (Noun-like Verb)

As illustrated in (3), there is little difference in sentence structure between the noun (1) and verb (2) items up until the word following the main verb of each sentence frame:

(3) NP V [the N]/[to V] ...

The main verbs were strongly biased to generate expectations for a NP for the noun items, and for an infinitival complement for the verb items (see Farmer et al., 2006 for information about these biases). Given the substantial amount of overlap in structure between the noun and verb items up until the NP or infinitival complement, it is possible that presenting the noun and verb materials together caused the biases of the main verbs to decrease, such that over the span of the experiment, expectancies for either a NP or infinitival complement were diminished. A consequence is that when the probability of encountering either a noun or a verb is diminished, as would be the case for the later items in the Staub et al. experiment, the effect of phonological typicality on reading times at the target word may be reduced or eliminated altogether. Hence, we would predict that interleaving items with different contextual influences would result in a gradual reduction of the phonological typicality effect.

To test this hypothesis, we followed Staub et al. in combining the original noun and verb items from Farmer et al.'s two separate experiments within a single self-paced reading experiment. If combining items that produce a strong expectation for a noun with the items that produce a strong expectation for a verb reduces the expectation for target words of either lexical category as the experiment progresses, we should make two observations:

1) When conducting the exact same linear mixed-effects analysis that Staub et al. report in their Experiment 2 (on self-paced reading), we should replicate their lack of a significant interaction between Part of Speech (PoS) and Phonological Classification (PC; whether the target word is Noun-like or Verb-like).

2) When adding Presentation Order to the model as a fixed effect, thus controlling for unique variance associated with the order in which the noun and verb items were viewed, we should see a PoS x PC interaction emerge, and should observe a PoS x PC x Order interaction. The phonological typicality effect—noun-like nouns being read faster than verb-like nouns in the noun context, and verb-like verbs being read faster in the verb context—should be present for the items each subject encountered early in the experiment, when the biases exerted by the main verbs remain strong. Later in the experiment, when expectations for either a noun or a verb have been attenuated, the typicality effect should weaken.

Method

Participants

Thirty-four undergraduate native English speakers from Cornell University (*M*=19.62 years, *SD*=1.13) participated for extra credit in a psychology course.

Materials

For both the noun and verb items, two sentence versions were constructed from each sentence frame. One version included a noun phrase with a noun-like noun (*marble,* 1a), and the other version contained a verb-like noun (*insect,* 1b). For the verb items, one version of each sentence frame contained an infinitival complement containing a verb-like verb (*assist*, 2a), and the other version contained a noun-like verb (*vary*, 2b). For both the noun and verb items, there was no significant difference in CELEX- and HAL-based lexical frequency, orthographic length, number of phonemes, number of phonological neighbors, or plausibility between the phonologically typical versus atypical items. The 20 experimental items (10 noun and 10 verb items) were combined and then counterbalanced across two different presentation lists in such a way that each list contained five noun-like noun sentences, five verb-like noun sentences, five verb-like verb sentences, and five noun-like verb sentences, but only one version of each of the 20 frames. Each list also contained 30 unrelated filler items and eight practice items. A majority of the filler sentences contained reduced or unreduced relative clauses, and the others were simple unambiguous sentences containing no relevant psycholinguistic manipulations.

Procedure

Subjects were randomly assigned to one of the two presentation lists. All sentences were randomly presented in a non-cumulative, word-by-word moving window format using PsyScope version 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993). After a brief tutorial, subjects were instructed to press the 'GO' key to begin the task. For all sentences, the entire test item appeared left-justified at the vertical center of the screen in such a way that dashes preserved the spatial layout of the sentence, but masked the actual characters of each word. As the subjects pressed the 'GO' key, the word that was just read disappeared and the next one appeared. RTs (msec) were recorded for each word. After each sentence had been read, subjects responded to a Yes/No comprehension question, and upon another key press, the next item appeared.

Results

One participant reported the presence of an auditory processing deficit and was excluded from all subsequent analyses. Overall accuracy on the comprehension questions relating to the 20 experimental sentences was close to ceiling (M=19.46 correct, SD=1.15), and no significant main effect of PC or PoS, or interaction, was observed on accuracy rates, all F's < 1.3. In keeping with the original Farmer et al. experiments, the focus of our analyses was on the critical word that contained the experimental manipulation of phonological typicality. All RTs over 2000msec were excluded from the subsequent analyses, resulting in the omission of five trials (less than 1% of the data).

The mean RTs on the critical word for each condition are presented in Figure 5.1. As in Staub et al., RTs on the critical word were analyzed in a linear mixed-

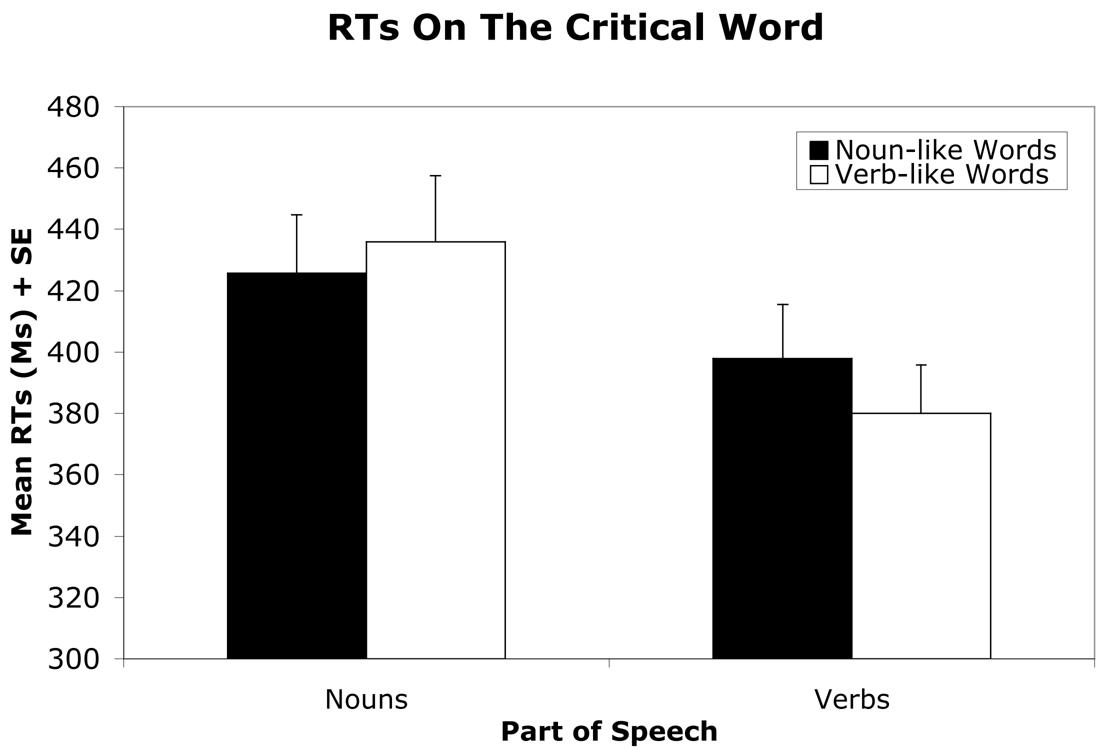

Figure 5.1. Mean RTs on the critical word for each condition of the PoS x PC interaction.

effects model using the lme4 package in R[9] (R Development Core Team, 2007), and the analyses will be presented twice, once without the inclusion of presentation order, and once with order as an additional fixed factor. In the analysis set without considering any effect of order, RTs were the dependent measure, subjects and items were entered as crossed random factors, and the fixed factors were PoS, PC, the PoS x PC interaction, length, and HAL-based log frequency. All parameter estimates, as well as *p*-values (estimated by Markov Chain Monte Carlo sampling, Baayen, 2008) associated with the t-tests for each effect, are listed in Table 5.1. As evident in Table 5.1, the results were similar to those of Staub et al. in that there was no significant

[9] We are grateful to Adrian Staub and Margaret Grant for making the R syntax for their statistical analyses available to us.

effect of PoS or PC, no significant interaction between PoS and PC, and no significant effect of frequency. Unlike Staub et al., however, there was a significant effect of length in the present dataset, with longer words being read more slowly.

Table 5.1

Regression weights for the mixed-effects model on critical-word RTs without assessing the effect of presentation Order.

	Estimate	p-value
Intercept	358.40	.0001
Part of Speech (PoS)	-12.90	.635
Phonological Classification (PC)	19.26	.394
PoS x PC	-44.36	.164
Length	18.81	.034
Log Frequency	-5.47	.394

In order to assess the hypothesis that the effect of phonological typicality would diminish as the experiment progressed, we conducted the same analysis detailed above, except that Presentation Order was entered as a fixed effect, interacting with PoS and PC. Table 5.2 displays the parameter estimates and p-values associated with each term in the model, and we focus our attention here on the interaction terms most relevant for illuminating the effect of Order. Crucially, after accounting for unique variance associated with Order, the PoS x PC interaction was significant, $p = .035$. Order did not interact with PoS, but it did with PC, such that the overall difference between noun-like and verb-like words decreased as the experiment progressed, regardless of PoS. Additionally, there was a marginally significant three-way interaction between Order, PoS, and PC, *p* = .082, indicating that the interaction between PoS and PC was dependent on Order.

Table 5.2

Regression weights for the mixed-effects model on critical-word RTs, taking into account the unique variance associated with presentation Order.

	Estimate	p-value
Intercept	389.50	.0001
Part of Speech (PoS)	37.02	.444
Phonological Classification (PC)	105.58	.028
PoS x PC	**-138.96**	**.035**
Length	19.06	.022
Log Frequency	-7.71	.246
Order	-1.08	.700
PoS x Order	-4.93	.200
PC x Order	-8.78	.025
PoS x PC x Order	9.66	.082

To illustrate the influence of presentation Order on the phonological typicality effect, bins of items were generated based on whether the items of each PoS condition appeared early or late in the experiment for each participant. More specifically, one bin contained the first five noun items, and another contained the last five noun items. Bins were also created for the first and last five verb items. Additionally, under the

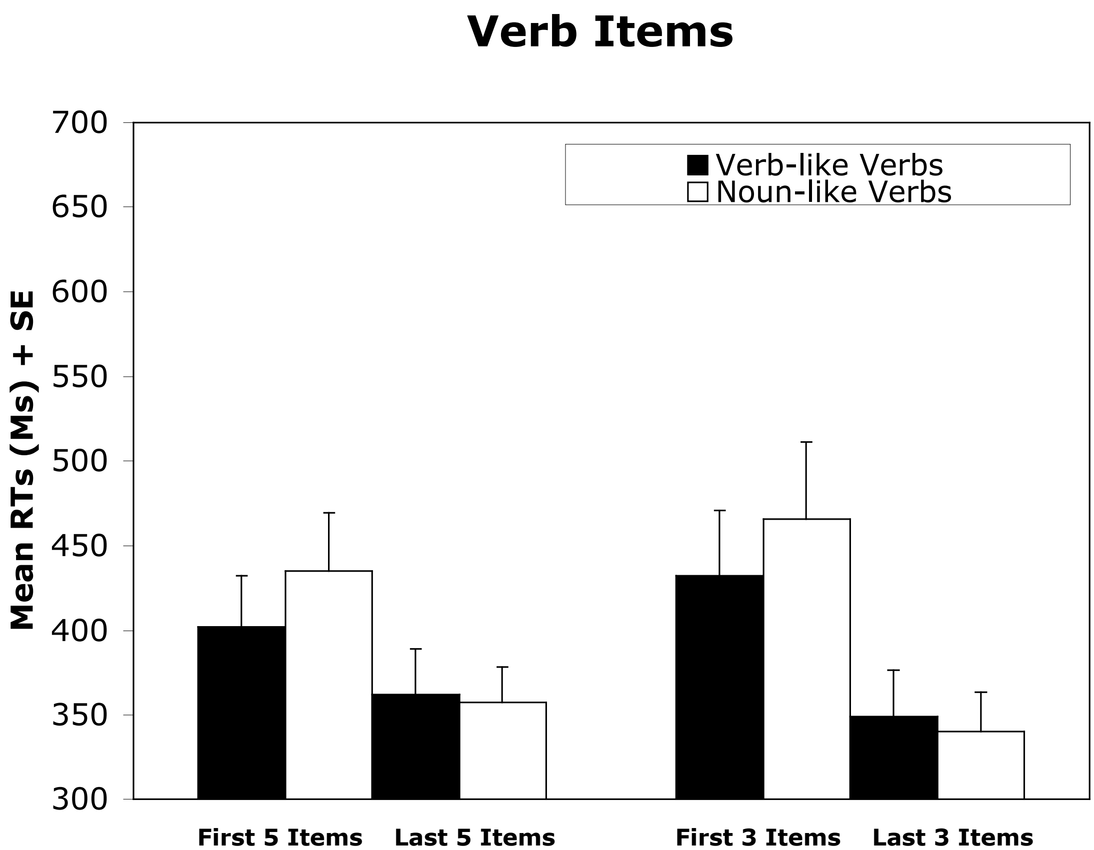

Figure 5.2. Average RTs across the first-and-last five (left) and three (right) verb items.

idea that as the experiment progressed, the syntactic expectancies for a NP or infinitival complement faded, thus diminishing the typicality effect, we also generated bins for the first and last three noun and verb items. Then, within both the early and late bins for each PoS, the magnitude of the typicality effect was assessed.

Figure 5.2 shows the predicted effect of Order. For both the first/last-five and the first/last-three verb items, verb-like verbs were read more quickly than noun-like verbs at the beginning of the experiment, but in the latter portion of the experiment, the effect of PC disappeared. As illustrated in Figure 5.3, there is a similar pattern for the noun items. The typicality effect existed, in the predicted direction, for the early items.

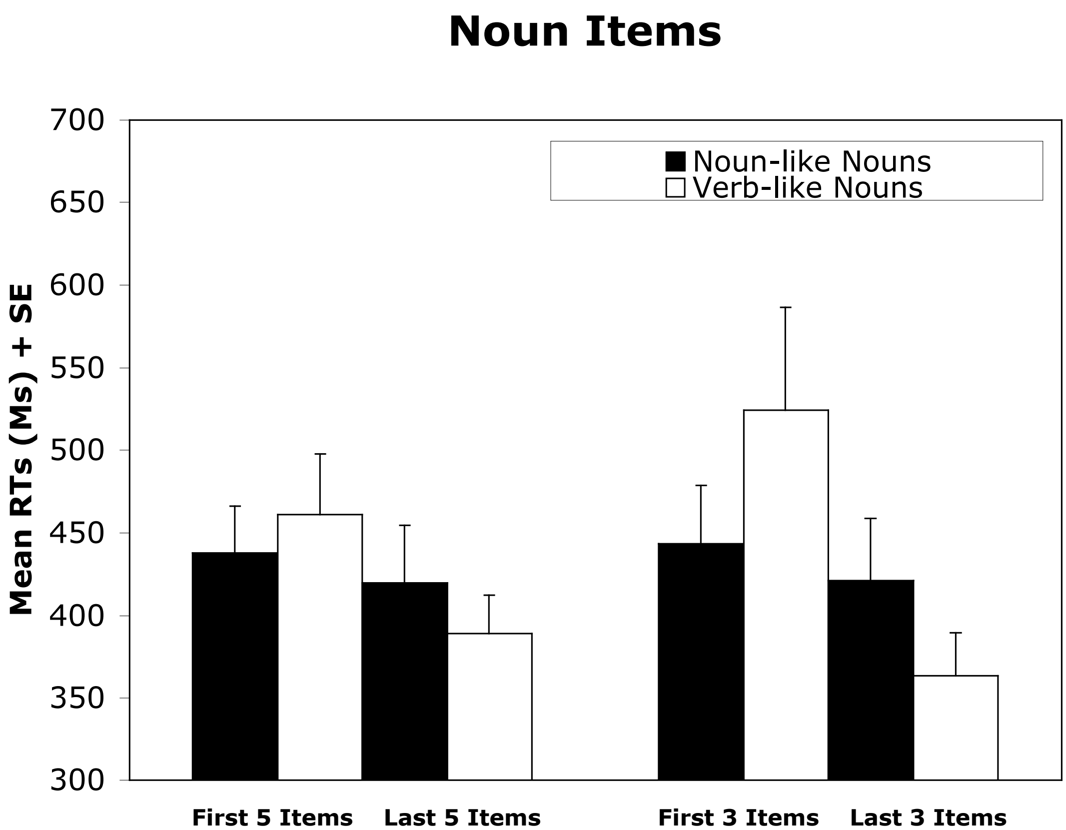

Figure5. 3. Average RTs across the first-and-last five (left) and three (right) noun items.

Interestingly, however, the typicality effect was larger for the first three items compared to the first five. The pattern of effects differ somewhat for the final noun and verbs items, suggesting that predictiveness of prior context may affect noun and verb phonological typicality in slightly different ways. In this case, context-driven expectancies appear to influence nouns more than verbs, perhaps because phonological typicality may be a stronger factor for verbs than for nouns (e.g., Christiansen & Monaghan, 2006; Fitneva et al., in press).

General Discussion

Based on their data, Staub et al. claimed that the phonological typicality effects reported in Experiments 2 and 3 of Farmer et al. were likely the result of a Type I error. The data presented here, however, provide support for the notion that their null results were likely due to their altering of the original Farmer et al. design by interleaving syntactic frames that generate a strong expectation for a noun with those that are highly predictive of verbs. Using their interleaved design, we found that after controlling for the unique variance associated with presentation Order, the critical PoS x PC interaction was significant, with the three-way interaction with Order being marginally significant. The effects of presentation Order observed here provide support for our hypothesis that the overlap in syntactic context preceding the critical words would reduce the strength of the expectation for either a noun or a verb over time, and that this reduction in main verb bias would have a negative impact on the typicality effect. As predicted by this hypothesis, we found that the typicality effect for each grammatical category decreased as the experiment progressed. For both the noun and verb items, the phonological typicality effect was observed for the items presented early, where main verb biases for either a NP or VP would be strongest, and was attenuated across the course of the experiment.

In their discussion, Staub et al. claim that should intermixing the noun and verb items cause the elimination of the phonological typicality effect, then the effect would "reflect task-dependent strategic factors as opposed to the processes involved in normal word recognition" (p. 813). In contrast, we contend that the data presented here reflect the fact that the predictiveness of a syntactic context can serve as one potential mediator of the phonological typicality effect. Reduction of contextual predictiveness was a consequence of the interleaved experimental design and thus cannot be taken as more representative of natural language processing. Note that participants began the study showing a strong effect of phonological typicality, which then decreased as the experiment progressed, likely as adaptation to the decrease in predictiveness of main verb biases in the stimulus set. Indeed, when controlling for presentation order, the phonological typicality effect remained reliable. However, we do not see these results as an end-point but rather as a launching pad for further experimental investigations into the relationship between phonological typicality, syntactic context, and other variables known to influence normal reading.

REFERENCES

Altmann, G. T. M. (1998). Ambiguity in sentence processing. *Trends in Cognitive Science, 2*, 145-152.

Altmann, G. T. M., van Nice, K. Y., Garnham, A., & Henstra, J.-A. (1998). Late closure in context. *Journal of Memory and Language, 38*, 459-484.

Baayen, R.H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Cassidy, K. W., & Kelly, M. H. (2001). Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin and Review, 8*, 519-523.

Christiansen, M.H. & Monaghan, P. (2006). Discovering verbs through multiple-cue integration. In K. Hirsh-Pasek & R.M. Golinkoff (Eds.), *Action meets words: How children learn verbs* (pp. 88-107). New York: Oxford University Press.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1992). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*(2), 257-271.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117-1121.

Elman, J. L., Hare, M., & McRae, K. (2004). Cues, constraints, and competition in sentence processing. In M. Tomasello & D. I. Slobin (Eds.), *Beyond nature-nurture: Essays in honor of Elizabeth Bates* (pp. 111-138). Mahwah, NJ: Erlbaum.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103*, 12203-12208.

Farmer, T. A., Monaghan, P., Misyak, J. B., & Christiansen, M. H. (*submitted*). Phonological typicality influences sentence processing in predictive contexts.

Fitneva, S. A., Christiansen, M. H., & Monaghan, P. (in press). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language.*

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37,* 58-93.

Hagoort, P. (in press). Reflections on the neurobiology of syntax. In D. Bickerton & E. Szathmáry (Eds.), *Biological foundations and origin of syntax. Strüngmann Forum Reports, Vol. 3*. Cambridge, MA: MIT Press.'

Hinojosa, J. A., Moreno, E. M., Casado, P., Muñoz, F., & Pozo, M. A. (2005). Syntactic expectancy: and event-related potentials study. *Neuroscience Letters, 378*, 34-39.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99*, 349-364.

Monaghan, P., Chater, N., & Christiansen, M. H. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology, 55*, 259-305.

Pickering. M.J. & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*, 105-110.

R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 720-732.

Snedeker, J., & Yuan, S. Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language, 58,* 574-608.

Staub, A., & Clifton, C., Jr. (2006). Syntactic prediction in language comprehension: Evidence from Either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 425-436.

Staub, A., Grant, M., Clifton, C., Jr., & Rayner, K. (2009). Phonological typicality does not influence fixation durations in normal reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 806-814.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443-467.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*, 1272-1288.

CHAPTER 6

Summary and Discussion

The results of the experiments presented in Chapters 2-5 provide evidence for a language comprehension system that operates by exploiting multiple sources of information, even phonological information, in pursuit of the ultimately correct interpretation of an incoming linguistic signal. In the face of ambiguity, competition among multiple simultaneously active representations occurs, and the competition dynamics are modulated by biases set forth by all salient and reliable sources of information available to the system. Below, after summarizing the results of the studies presented here, they will be discussed in the context of the theoretical frameworks that they advance.

The data presented in Chapter 2 serve as an initial demonstration that mouse-movements are capable of indexing the garden-path and context effects that have been demonstrated in previous eye-movement studies (e. g. Snedeker & Trueswell, 2004; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In a one-referent visual context where one location corresponds to the garden-path interpretation of a spoken sentence (akin to a distractor), significant curvature toward the distractor occurs as the correct object is moved to the location that corresponds to the ultimately correct interpretation of the sentence. Although many other studies have now demonstrated that mouse movements are indicative of the competition process between a target and distractor as it unfolds over time, the degree to which movement trajectories could capture competition dynamics associated with temporarily ambiguous sentences was less clear. Most of the other studies that have used movement trajectories to study processing in other domains have dealt with a two-choice display in which there was either no temporal ambiguity at all (e. g. Dale, Kehoe, & Spivey, 2007; Freeman, Ambady, Rule, & Johnson, 2009; Song & Nakayama, 2008), or a temporal ambiguity that was only present for an extremely short amount of time (e. g. Farmer & Zevin, 2008; Spivey, Grosjean, & Knoblich, 2005). In the visual-world task used here, the scene was more complex, containing 5-6 objects, and the duration of the ambiguity was much longer (spanning three words instead of approximately 1-5 phonemes) than in previous studies. As such, eliciting trajectory curvature toward a distractor in this more complex task, and replicating the context and garden-path effects typically studied by examining patterns of saccades around a visual scene, provides support for the reliability of data produced by the technique.

Although the layout of the display used in the preliminary study detailed in Chapter 2 was slightly problematic due to the fact that velocity and spatial attraction were confounded, the problem was corrected in the first study of Chapter 3, and the garden-path and context effects again replicated. A much more thorough analysis of the shapes of the distributions across all conditions was conducted, and maximum deviation values were used as a trial-by-trial index of garden-path magnitude instead of area-under-the-curve values. Once again, no evidence of bimodality was detected in the garden-path condition, but instead, the response distribution appeared continuous / unimodal. As outlined in Chapter 1, these distributional analyses are problematic for the unrestricted race account, but provide further support for a multiple constraint-based account of the data. Moreover, the competition-integration simulation detailed in the second study of Chapter 3 produces a unimodal distribution of responses in the garden-path condition. This simulation goes some way in terms of validating the claim that a competition-based account of garden-path resolution actually does predict a continuous distribution of garden-path effects.

In relation to the final study of Chapter 3, proponents of continuous dependent variables, such as arm-movement trajectories, have argued that cognitive processes that are fundamentally continuous in nature are masked by the discrete nature of dependent variables typically used in behavioral research (Spivey, 2007; Spivey, Grosjean, & Knoblich, 2005). However, given the continuous nature of arm-movement data, one might be compelled to argue the opposite. That is, it could be the case that the continuous nature of arm-movement trajectories is actually masking a discrete process. Study three was conducted in order to determine whether a bimodal distribution of responses would be detectable in a situation where it would be highly expected to occur, and indeed, it was.

The four studies contained within Chapters 2 and 3 provide relatively strong support for a competition-based versus unrestricted race account of the data. Two highly inter-related issues, however, deserve further attention. First, there is some issue about just how on-line the trajectories in this paradigm are. Indeed, in Chapter 2, raw-time analyses were conducted, and showed that the divergence effects were occurring within roughly the same time range as they were in the eye-movement versions of the experiment. Although that is true, it is also important to note that approximately 35-40% of the trajectories had to be excluded because they were initiated long (at least 2000 milliseconds) after the sentence had stopped playing. In Chapter 3, roughly the same trend occurred. Most of the trajectories had been initiated within 2000 milliseconds after the sentence stopped, but only about 75% were initiated while the sentence was still being heard.

This delay in initiation time may have ramifications for results reported in the two mouse-movements experiments. In the distribution of trajectories in the garden-path condition, Figure 2.4, for example, although there were a series of trajectories that exhibited intermediate curvature toward the incorrect destination, a large number of trajectories exhibited only very minor curvature toward the distractor, and some even curved away from it. It could be the case that those less-curved trajectories are the ones that were initiated much later than the trajectories exhibiting noticeable curvature toward the distractor. Indeed, an analysis not reported in either Chapter 2 or 3 sheds some light on this issue. It turns out that there is slight, albeit non-significant, negative correlation between the point in which a trajectory was initiated and the magnitude of the area-under-the-curve values in that dataset, r=-.17. Trajectories initiated earlier tended to exhibit a larger amount of trajectory curvature than did later-initiated trajectories.

Arm movements do take longer to initiate than saccades, and in relation to the data presented here, it is not immediately clear that the increased initiation time does not decrease the probability that any single trajectory will actually pick up on the dynamics of the competition process. One way to address this concern is to track both eye- and hand-movements at the same time. As outlined in Chapters 2 and 3, distributions of saccades are inherently bimodal, and as such, cannot readily provide information that can discriminate between a constraint-based versus unrestricted race account of syntactic processing. The pattern of saccades around the display will provide better information regarding the timing of fixations to the incorrect destination (typically interpreted as evidence for garden-pathing), and the timing between saccades and trajectory curvature can be explored in order to more fully understand the differential temporal sensitivity of each dependent measure. The two measures, together, can certainly illuminate both the timing and the dynamics of the competition process much better than can either measure by itself.

The data presented in Chapters 2 and 3 were used, by and large, to aid in the process of discriminating between the unrestricted race versus constraint-based accounts of syntactic ambiguity resolution. However, the question about what these data contribute to debates between the syntax-first and constraint-based models deserves some attention. In true spirit of the field of syntactic processing, a relatively “acrimonious” debate has ensued between the syntax-first staged-based and the interactive dynamic accounts of syntactic processing. Although some stage-based accounts allow for a very limited interaction between syntactic and non-syntactic information sources during the first stage of processing (e. g. Abney, 1989; Crain & Steedman, 1985), most of these accounts still attribute initial analysis only to automatically-applied syntactic heuristics, without regard to other present and potentially useful cues. As noted in Chapter 1, however, many of the constraint-based papers cited above have demonstrated that factors such as referential context, thematic fit, and prosody do seem to facilitate avoidance of the garden-path effect, and that the effect of those information sources is, in many cases, as immediate as possibly detectable by behavioral testing methods. Results such as these tend to bolster accounts of syntactic processing that are not as informationally encapsulated as the hard-lined versions of a stage-based syntax-first model, and are indeed problematic for those accounts.

The influence of non-syntactic information on early processing behavior is, however, by no means conclusive evidence against traditional syntax-first models. Proponents of the syntax-first models argue that behavioral testing methods are not temporally sensitive enough to accurately index initial parsing (e. g. Clifton & Ferreira, 1989). Instead, they typically argue that the observed effects of non-syntactic information can be attributed to the reanalysis mechanism that, as explained above, is hypothesized to be sensitive to many different information sources, instead of the comprehension machinery necessary for initial structural decisions (e. g. Fodor & Ferreira, 1998; Friederici, 1995; 1998; 2002). Although an influence of scene-based referential context was detected in the studies detailed in Chapters 2 and 3, given the delayed initiation of mouse-movements, proponents of syntax-first models would almost certainly attribute it to the reanalysis mechanism.

In relation to the temporal sensitivity of dependent measures, one would assume that measures as real-time as an electroencephalogram would aid in adjudicating the debate between syntax-first versus interactive processing. However, not even ERP studies investigating the influence of non-syntactic information on early processing have not even been able to provide definitive evidence in support of one side over the other. It is true that the presence of a P600, typically thought to index some type of syntactic expectancy violation, can be modulated by the presence or absence of, say, extra-sentential contextual information (Brown, van Berkum, & Hagoort, 2000; van Berkum, Brown, & Hagoort, 1999; van Berkum, Brown, Hagoort, & Zwitserlood, 2003). In studies such as these, a P600 occurs when a participant is garden-pathed, and can be eliminated by non-syntactic cues that support the initially less-active syntactic alternative.

However, there is considerable debate about what the P600 actually indexes (e. g. Dikker, Rabagliati, & Pylkkanen, 2009; Friederici, 1998; Frisch, Schlesewsky, Saddy, and Alpermann, 2002; Kaan, Harris, Gibson, and Holcomb, 2000). For example, many studies have reported, in addition to the P600, syntax-elicited negative going waveforms with an anterior, left-lateralized scalp distribution and an early, albeit variable, post-stimulus onset time (Friederici, Hahne, & Mecklinger, 1996; Gunter, Friederici, & Schriefers, 2000; Hahne & Friederici, 1999). One of the most well known of these "early" syntax components is the the Early Left Anterior Negativity (ELAN). The ELAN has a post-stimulus onset of ~ 100 ms and typically occurs only in relation to phrase structure violations brought about by the misplacement of a closed-class "function" word (Hahne & Friederici, 1999), such as when a preposition is encountered after a determiner, etc (but definitely see Dikker, Rabagliati, & Pylkkanen, 2009; Dikker, Rabagliati, Farmer, & Pylkkanen, *submitted*). As such, proponents of syntax-first models argue that the "parser" is in charge of demarcating the structure of a sentence, typically by way of grammatical category assignment, and that those processes are indexed by the ELAN. The ramification for the P600, then, is that it is not indicative of the initial-stage of parsing, but instead is an indicator of the engagement of the stage two reanalysis mechanism (Hahne & Friederici, 1999; Friederici, 2002).

So, the presence of the context effect in the two referent condition does not help rule out a syntax-first model due to the non-immediacy of the mouse-movement data. But, what about the unimodal distribution of garden-path effects? In very hard-lined versions of a syntax-first model where no noise is built into the system, one would expect that syntactic heuristics would always select the incorrect interpretation of the ambiguity first, such that reanalysis would always be needed. In terms of response distributions, then, hard-lined syntax first models would predict a very tight unimodal distribution of responses with all trials corresponding to an immediate commitment to the wrong interpretation, characterized by deliberate movement toward the incorrect destination, followed by a redirection toward the correct location.

In terms of the mouse-movement data, then, syntax-first models would predict distributions of movement similar in nature to the "switch condition" in experiment 3 of Chapter 3, where a discrete initially-incorrect movement is made, followed by a corrective movement aimed toward the ultimately correct location. In Figures 2.4 and and 3.3, however, we see that only a very few trials exhibited extreme movement toward the incorrect destination, before a redirection occurs. This pattern of results is unlike the pattern of "switch" trials in experiment 3 of Chpater 3 (Figure 3.10). This fact is somewhat problematic for syntax-first accounts, although yet again, the gradiency observed in the garden-path condition can be readily attributed to a reanalysis mechanism. It may be that the "repair" end of hard-lined syntax-first models is susceptible to gradiency produced by being influenced by many different factors that aid in the repair process. As such, the non-immediacy of the mouse-movement makes it difficult to address the debate between syntax-first and constraint-based models, although the presence of data that trends away from the predictions of syntax-first models is, it is at the very least, promising. Hopefully, combining more temporally sensitive measures of cognitive processing with the continuous mouse-movement data will, in the future, finally lay that debate to rest.

Phonological Typicality

Interpretation of the results contained in Chapters 4 and 5 is, in many respects, much more straightforward. The reliable, coherent clustering of nouns and verbs in phonological space, based only on their phonemic contents, demonstrates that differential statistically reliable phonemic information is present in words of each grammatical category. Moreover, the fact that a quantitative measure of phonological typicality was able to account for unique variance in word naming latency provides preliminary evidence that people are sensitive to those phonological regularities. These facts, in and of themselves, are problematic for the age-old assumption that there is no reliable mapping between the form of a word and its meaning (de Saussure, 1916). That is, to the degree that the noun/verb word class distinction can been seen as a rudimentary form of meaning, the coherent clustering of nouns and verbs proves problematic for hard-lined versions of the arbitrariness claim. Instead, the results align well with a series of more recent studies that have demonstrated other links between various properties of word form and meaning (e. g. Arciuli & Monaghan, 2009; Bergen, 2004; Nuckolls, 1999; Nygaard, Herold, & Namy, 2009).

These types of systematic relationships between word form and meaning are not just another interesting tidbit about language. Instead, the studies in Chapters 4 and 5 provide evidence that the phonological typicality variable has very real ramifications for the processes underlying both normal reading and syntactic processing. In studies 2 and 3 of Chapter 4, it is demonstrated that when a sentential context sets up a very strong expectation for either a noun or a verb, reading times are slower when the phonological properties of the noun (when expecting a noun) or verb (when expecting a verb) are atypical of other words in their grammatical category. As noted in Chapter 5, Staub, Grant, Clifton, and Rayner (2009) have failed to replicate these two effects. Although in Farmer, Christiansen, & Monaghan (2006) each experiment was run separately, Staub et al. combined the noun and verb items, and failed to replicate the Farmer et al. results in both an eye-movement study and a self-paced reading study. It should be noted, however, that combining the noun and verb items into one study represents a huge departure form the original design. As demonstrated in Chapter 5, when the two sets of items are combined, the phonological typicality effect on normal reading disappears. As argued in Chapter 5, the loss of the effect is more than likely due, however, to a reduction in the expectation for either a noun or a verb that occurs progressively across the study, as demonstrated by the effect of Order detailed in Chapter 5.

More recently, the effects of phonological typicality on normal reading have been replicated by a (relatively) independent group using a novel paradigm. Indeed, as noted by Tanenhaus and Hare (2007), studies of reading have found that initial saccades to words are relatively uninfluenced by various types of linguistic information that typically exert an influence on later processing. They argued that during reading, predictions about upcoming word forms are being made, and that various cues to word form, such as phonological typicality, may be the types of factors that would influence initial saccades to upcoming words. Although this hypothesis was not directly tested by Tanenhaus and Hare, it was, indirectly, by Dikker, Rabagliati, Farmer, & Pylkkanen, *submitted*). Using magnetoencephalography (MEG), Dikker et al. demonstrate that the M100, a component in visual cortex that arises approximately 100 milliseconds after stimulus onset in response to violations of expectedness, is sensitive to phonological typicality. They found that an effect of expectedness of a noun (should a noun be next or not) was modulated by the phonological typicality of the incoming noun. In a condition where all nouns had phonological properties highly typical of nouns, the effect of expectedness was larger than in a condition where all of the nouns were neutral in terms of their phonology. That is, the M100 effect was significantly larger when a noun was not expected but occurred and was highly typical of other nouns, than when a noun was expected. When the nouns were not typical or atypical of other nouns (neutral), there was no difference in the expected versus the unexpected condition. Interestingly, this effect appears to be generated in the visual cortex while reading, and is in-line with the Tanenhaus and Hare proposal (also advanced in Dikker, Rabagliati, Farmer, & Pylkkanen, *submitted*) that while reading, word form predictions of upcoming material are being generated. This effect happens so early that it may not be detectable in eye-movement data. Nonetheless, it accentuates the role that word form predictions may exert during language processing.

The final study in Chapter 4 shows that not only does phonological typicality influence the processing of single words in simple unambiguous sentences, but that it can also influence the interpretation of an ambiguity. In sentences where a temporary structural ambiguity arises as a result of a noun / verb homonym, the interpretation of that homonym is biased toward “noun” or “verb” based on whether it has phonological properties that are highly typical of a noun or of a verb. When the homonym is noun-like, and the ambiguity is resolved in accordance with the verb phrase interpretation of the ambiguity, a garden-path effect emerges. When the homonym is verb-like, and yet the ambiguity is resolved in accordance with the noun compound interpretation of the ambiguity, a garden-path effect also arises.

In terms of models of on-line syntactic processing, the effect of phonological typicality in this study seems to bolster a conatraint-based account of the data, whereby all salient and reliable sources of information can exert an influence on processing. Indeed, the effect of phonological typicality is the first demonstration that a variable based only on the phonemic properties of a word can also exert an influence on the processing of garden-path sentences. Although constraint-based accounts can most certainly accommodate the phonological typicality effect, it is unclear whether or not this effect is truly problematic for a syntax-first model. First, the data were collected using self-paced reading, which provides, at most, only 1-2 data points per second. As a result, the temporal time-frame in which this variable actually exerts its influence is not well established (although, as noted directly above, the MEG data of Dikker et al. suggest that the phonological typicality effect occurs extremely early in processing). So of course, proponents of syntax first models could feasibly argue that phonological typicality is another variable that influences reanalysis instead of first-stage processing. It is worth pointing out, however, that the coherent clustering of nouns and verbs in phonological space implies an implicit correlation between phonological form and word category, such that the two may not be distinguishable. To that end, an effect of phonological typicality on first-stage parsing under a syntax-first account, where word category assignment is one of the key functions of the syntactic parsing mechanism (e. g. Frazier & Fodor, 1978; Friderici, 2002), may be readily predicted.

Conclusion

The results of the studies presented in the pages of this volume serve to address fundamental questions about the architectures and mechanisms responsible for fast, accurate, and effortless language comprehension. Building on the multiple-constraint based models of sentence processing advanced in the 1990s, these results demonstrate the graded nature of the garden-path effect, and show that a competition-based mechanism of syntactic ambiguity resolution can account for the observed gradedness where other classes of models fail. Moreover, they show that even the often-overlooked variable of phonology can serve to facilitate the comprehension process. Indeed, this collection of experiments lays the foundation for future work that can further help explain the complex processes that give rise to various aspects of verbal and written communication.

REFERENCES

Abney, S. P. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research, 18*, 129-144.

Arciuli, J., & Monaghan, P. (2009). Probabilistic cues to grammatical category in English orthography and their influence during reading. *Scientific Studies of Reading, 13*, 73-93.

Bergen, B. K. (2004). The psychological reality of phonaestemes. *Language, 80*, 290-311.

Brown, C. M., van Berkum, J. J. A., & Hagoort, P. (2000). Discourse before gender: An event-related potential study on the interplay of semantic and syntactic information during spoken language understanding. *Journal of Psycholinguistic Research, 29*, 53-68.

Clifton, C., & Ferreira, F. (1989). Ambiguity in context. *Language and Cognitive Processes, 4*, 77-103.

Crain. S., & Steedman, M. J. (1985). On not being led up the garden-path: The use of context by the psychological processor. In D. Dowty, L. Kartunnen, & A. Zwicky (Eds.), *Ntural language parsing* (pp. 320-358). Cambridge, UK: Cambridge University Press.

Dale, R. A., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in time course of categorizing atypical exemplars. *Memory and Cognition, 35*, 15-28.

de Saussure, F. (1916) *Cours de Linguistique Generale*. Paris, France: Paycot.

Dikker, S., Rabagliati, H., & Pylkkanen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition, 110*, 293-321.

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkanen, L. (*submitted*). Early occipital sensitivity to syntactic category is based on form typicality.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103*, 12203-12208.

Farmer, T. A., Liu, R., Mehta, N. S., & Zevin, J. D. (2009). Native language experience influences the perceived similarity of second language vowel categories. In *Proceedings of the Cognitive Science Society*.

Farmer, T. A., Monaghan, P., Misyak, J. B., & Christiansen, M. H. (*submitted*). Phonological typicality influences sentence processing in predictive contexts.

Fodor, J. D., & Ferreira, F. (Eds.) (1998). *Reanalysis in sentence processing*. Netherlands: Kluwer.

Freeman, J. B., Ambady, N., Rule, N., O., & Johnson, K. J. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General, 137*, 673-690.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological data. *Brain and Language, 50*, 259-281.

Friederici, A. D. (1998). Diagnosis and reanalysis: Two processing aspects the brain may differentiate. . In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 177-200). Netherlands: Kluwer.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*, 78-84.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1219-1248.

Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The p600 as an indicator of syntactic ambiguity. *Cognition, 85*, B83-B92.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12*, 556-568.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*, 194-205.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The p600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*, 159-201.

Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology, 28*, 225-252.

Nygaard, L. C., Harold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science, 33*, 127-146.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49*, 238-299.
Psychology: Learning, Memory, and Cognition, 24, 1521-1543.

Song, J., & Nakayama, K. (2008). Numeric comparison in a visually-guided manual reaching task. *Cognition, 106*, 994-1003.

Spivey (2007). *The continuity of mind*. NY, NY: Oxford University Press.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proc. National Academy of Sciences, 29*, 10393-10398.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of syntactic context on syntactic ambiguity resolution. *Cognitive Psychology, 45*, 447-481.

Staub, A., Grant, M., Clifton, C., Jr., & Rayner, K. (2009). Phonological typicality does not influence fixation durations in normal reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 806-814.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*, 147-182.

van Berkum, J. J. A., Brown, C. M., & Hagoort, P., & Zwitserlood, P. (2003). Event-related potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology, 40*, 235-248.